

AD-A205 192

## REPORT DOCUMENTATION PAGE

2

Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		Approved for public release; distribution unlimited.	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
University of Rochester Laboratory for Laser Energetics		ARO 24481.1-EL-SDI	
6a. NAME OF PERFORMING ORGANIZATION		6b. OFFICE SYMBOL (If applicable)	
U. S. Army Research Office		7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code)		7b. ADDRESS (City, State, and ZIP Code)	
200 East River Road Rochester, NY 14623-1299		P. O. Box 12211 Research Triangle Park, NC 27709-2211	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)	
U. S. Army Research Office		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code)		DANL03-86-K-0152	
P. O. Box 12211 Research Triangle Park, NC 27709-2211		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO	PROJECT NO
		TASK NO	WORK UNIT ACCESSION NO
11. TITLE (Include Security Classification) Time-Resolved Hot Electron Transport in Electronic Devices			
12. PERSONAL AUTHOR(S) G. Mourou, J. Whitaker, T. Morris, and K. Meyer			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 9/30/86 TO 8/31/88		14. DATE OF REPORT (Year, Month, Day) December 1988
16. SUPPLEMENTARY NOTATION		The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.	
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) hot electron transport, velocity overshoot, resonant tunneling diode, quantum well tunneling, time-resolved photoluminescence, subpicosecond absorption spectroscopy	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
The size of microelectronic devices is becoming increasingly small and they are operating at higher and higher frequencies up to and exceeding 100 GHz. Their performance and operation depends to a large extent on their carrier transport. Due to fast time variations and strong spatial nonuniformities, the electrons can not be considered classical anymore. The basic Einstein and Nyquist relations are not valid, the transport is not ohmic and the carriers are called hot electrons. The study of the hot-electron transport is of fundamental importance since it is prevailing in modern devices. It was the goal of the contract to directly study by means of femtosecond techniques the nonsteady-state transport of hot electrons in bulk or layered semiconductors.			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> CLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS PPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

19. (continued)

During the course of this contract we have developed a novel laser system producing 100-fs pulses at the microjoule level at kilohertz repetition rates. At this energy level a light continuum can be produced. This laser system allowed us to study both in absorption and luminescence the thermalization of the hot-electron population when subjected to a strong electric field, both for bulk and layered materials. We have also used the electro-optic sampling technique to obtain the electrical signature of fast devices, such as photoconductive gaps and resonant tunneling diodes, in the subpicosecond regime.

We have studied velocity overshoot in GaAs. This phenomenon, predicted by Ruch, has been observed for the first time by our group, by time-resolving photoconductivity in the first picosecond. The time-resolved photoconductivity exhibits a strong overshoot when the carriers are excited close to the energy gap. As the excess energy increases the overshoot effect is diminished, and there is also a delay in the onset of the rise time of the photocurrent. This delay, known as the Jones-Rees effect, is due to an excess carrier population with negative velocity. The dynamics of the hot-electron distribution as a function of electric field has been studied by time-resolved absorption techniques.

An extensive study of tunneling for single quantum wells as well as asymmetric quantum wells was performed as a function of electric field. The escape time was measured by monitoring the decay of the luminescence with a streak camera. This time-resolved study revealed a blue shift in the luminescence contrary to previous observations. This blue shift could be caused by a dipole field developing between the wells. Another explanation could be that the holes left behind could make one of the wells p-doped. A third explanation could be a reduction of the exciton binding energy by free carriers screening effect. In the case of asymmetric wells a charge transfer state has been observed. As the electric field is increased the charge state is strongly shifted to lower energy. The time-resolved photoluminescence revealed a long lifetime exceeding 10 ns.

Finally, a large amount of work has been performed to measure the switching time of the resonant tunneling diode. A rise time of 1.9 ps was measured. This represents the first experimental observation of picosecond bistable operation in a double barrier diode. It is the fastest switching event yet observed for an electronic device. Further measurements at different temperatures need to be made to differentiate between sequential and resonant tunneling.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

# Time-Resolved Hot Electron Transport in Electronic Devices

## Final Report

By G. Mourou, J. Whitaker, T. Norris and K. Meyer

December 1988

U. S. Army Research Office

Contract No. DAAL03-86-K0152

University of Rochester  
Laboratory for Laser Energetics  
250 East River Road  
Rochester, NY 14623-1299

Approved for public release;  
distribution unlimited

Accession For	
NTIS	CRA&I <input checked="" type="checkbox"/>
DTIC	TAB <input type="checkbox"/>
Unannounced <input type="checkbox"/>	
Justification _____	
By _____	
Distribution / _____	
Availability Codes	
Dist	Avail. J. or Special
A-1	

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE  
THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL  
DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO  
DESIGNATED BY OTHER DOCUMENTATION.

## Table of Contents

I.	<b>The Ultrafast Investigation Techniques.</b>	3
	In this report we are describing the optical and electrical techniques that we had to implement to directly investigate the electron transport in the picosecond and femtosecond time domain.	
I.A.	<b>Time-Resolved Photoluminescence</b>	3
I.A.1	Laser Oscillator	3
I.A.2	Time-Integrated Photoluminescence Spectroscopy	4
I.A.3	Time-Resolved Photoluminescence Spectroscopy	5
I.A.4	Cryogenics	7
I.B.	<b>Subpicosecond Absorption Spectroscopy</b>	7
I.B.1	The Dye Laser	8
I.B.2	Nd:YAG Regenerative Amplifier	12
I.B.3	Dye Amplifier	15
I.B.4	Continuum Generation and Amplification	16
I.B.5	Pump-Probe Experiments	19
I.B.6	Data Acquisition	24
I.C.	<b>Description of the Electro-Optic Sampling Technique</b>	26
II.	<b>Velocity Overshoot in GaAs</b>	31
II.A.	<b>Description of the Two Approaches: Time-Resolved Photoconductivity and Transient Absorption Spectroscopy</b>	31
II.B.	<b>Time-Resolved Photoconductivity</b>	33
II.B.1	Prediction of the Monte Carlo Theory	33
II.B.2	Experimental Observation of Velocity Overshoot and the Jones-Rees Effect	38
II.B.2.1	Excitation at 620 nm	39
II.B.2.2	Excitation at 720 nm	45
II.C.	<b>Transient Absorption Spectroscopy</b>	47
II.C.1	Monte Carlo Model of the Dynamic Electron Distribution Function	47
II.C.2	Experimental Results: Subpicosecond Heating and Thermalization of the Electron Distribution Function	54
II.C.2.1	Continuum Probe Results	54
II.C.2.2	Discrete Probe Results	59

<b>III. Tunneling in Quantum Wells</b>	<b>67</b>
<b>III.A. Tunneling from a Single Quantum Well</b>	<b>67</b>
III.A.1. Experimental Results	70
III.A.1.1 Sample Structure	70
III.A.1.2 Excitation and Luminescence	72
III.A.1.3 Photoluminescence Decay Times	73
III.A.1.4 Field Dependence of Photoluminescence Decay	77
III.A.1.5 Photoluminescence Stark Shifts	80
<b>III.B. Theoretical Interpretations</b>	<b>81</b>
III.B.1. Electric Field vs Bias Voltage	81
III.B.2. Tunneling-Time at Zero Field	82
III.B.3. Field Dependence of the Tunneling Time	83
III.B.4. Luminescence Blue Shift	85
<b>III.C. Tunneling from Asymmetric Quantum Wells</b>	<b>86</b>
III.C.1. Photoluminescence Experiments	86
III.C.1.1. Sample Design	87
III.C.1.2. Electronic States	88
III.C.1.3. Experimental Set Up	90
III.C.1.4. Experimental Results	91
<b>IV. Time-Resolved Resonant Tunneling</b>	<b>114</b>
<b>IV.A Theoretical Consideration</b>	<b>114</b>
IV.A.1. Current Voltage Characteristics	117
IV.A.2. Coherent vs Sequential Tunneling	119
<b>IV.B. Tunnel Diode Operation</b>	<b>122</b>
<b>IV.C. Switching Time Measurement</b>	<b>126</b>
IV.C.1. Test Experiment	128
IV.C.2. Experimental Observation	133
<b>REFERENCES</b>	<b>139</b>

## I. The Ultrafast Investigation Techniques

### I.A. Time-Resolved Photoluminescence

#### I.A.1. Laser Oscillator

The dye laser used in the time-resolved photoluminescence (PL) experiments was of a standard design. The pump source was a cw mode-locked Nd:YAG laser (Quantronix model 116, mounted on a Super-Invar slab for high thermal and mechanical stability). The Nd:YAG laser was acousto-optically mode-locked with a 50 MHz rf source, and the laser repetition rate was 100 MHz. Typical output parameters were 7 Watt average power, 80-ps pulse width, and 1% peak-to-peak power fluctuations. The output was frequency-doubled in a 5-mm KTP crystal, giving an average green power of 0.8 N-1.0 W. As was first shown by Sizer, *et al.*,<sup>1</sup> the frequency-doubled cw mode-locked Nd:YAG laser is a nearly ideal pump source for picosecond dye lasers because of its high stability, short-pulse width, and, as I will discuss later, it makes possible high gain synchronous amplification of short pulses.

The dye laser cavity was a standard folded astigmatically compensated design<sup>2</sup> with cavity length equal to that of the Nd:YAG pump laser, so the repetition rate was also 100 MHz. The laser setup is shown in Fig. 1. Laser dyes LDS 721 or LDS 698 (Pyridine 1) were used depending on the desired pump wavelength. The laser was tuned with a single-plate birefringent filter (Lyot) and (if necessary to force the laser to operate at a single wavelength near the edge of its tuning range), a 5  $\mu$ m uncoated pellicle. The output power and pulse width depended, of course, on the laser wavelength and tuning elements in the cavity. What is important for the PL experiments, however, is that the pulse width was always shorter than the 20-ps time resolution of the streak camera, and the output powers generally had to be kept lower than about 60 mW to avoid the generation of satellite pulses.

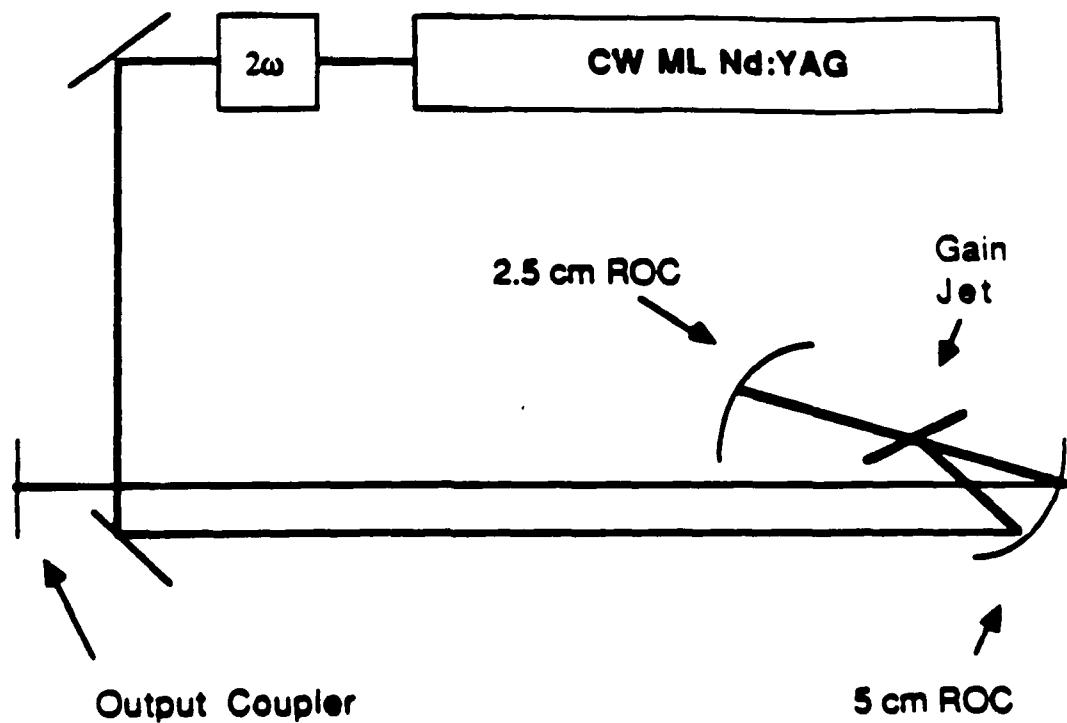


Figure 1: Synchronously pumped near-infrared dye laser used for time-resolved photoluminescence studies.

### I.A.2 Time-Integrated PL Spectroscopy

Luminescence from the sample was collected by an f3 lens. For time-integrated (cw) PL spectroscopy, the luminescence was imaged through a 0.32 m grating monochromator with 300 1-mm grating (Instruments SA model HR320) onto an optical multichannel analyzer (OMA). The OMA detector head was an EG&G PAR model 1420R, which is an intensified 1024-element diode array with extended red sensitivity. The OMA was controlled using an EG&G model 1461 controller interfaced to the PDP-11 computer via a direct-memory-access (DMA) board. A shutter in the pump beam was used to enable subtraction of background noise from the spectrum. The

spectral resolution of the system was 7 Å (1.3 meV), and the spectral range covered on a single spectrum was 2160 Å (365 meV).

### I.A.3 Time-Resolved PL Spectroscopy

The PL spectroscopy system is shown in Fig. 2. A mirror on a kinematic mount was used so that the PL could be directed to either the OMA for cw spectroscopy, or the streak camera for time-resolved studies. The streak camera was a Hamamatsu model C1587 with model M1955 synchroscan drive unit. The output of a PIN photodiode monitoring the output of the Nd:YAG laser oscillator was amplified, and this 100 MHz rf was used to drive the deflection plates of the synchroscan streak camera. Thus the PL signal arriving at the streak camera photocathode at 100 MHz was synchronized to the voltage driving the deflection plates (hence the term "synchroscan"). The jitter between the deflection plate voltage and the dye laser output is the principal limitation on the time resolution of the streak camera system. For typical signal integration times (1 second to a few minutes) this jitter limits the time resolution to about 20 ps.

The photocathode was an extended-red S-1 type, which covers the spectral region 300-1500 nm. The streak camera images the output of the photocathode onto a detector, and this imaging capability was used to enable us to do time-resolved spectroscopy. The luminescence from the sample was dispersed through a 0.32 m monochromator with 300 1/mm gratings across the entrance slit of the streak camera. A spectral resolution of about 3 meV could be achieved, with a total coverage of about 110 meV on a single spectrum. The streaked image was integrated on a two-dimensional intensified SIT detector (Hamamatsu model C1000) and stored on floppy disk for later analysis. It should be noted that the temporal dispersion of the luminescence signal by the grating was only about 0.5 ps over the entire detected spectrum, so no correction of the spectra for the group-velocity dispersion of the grating was necessary.

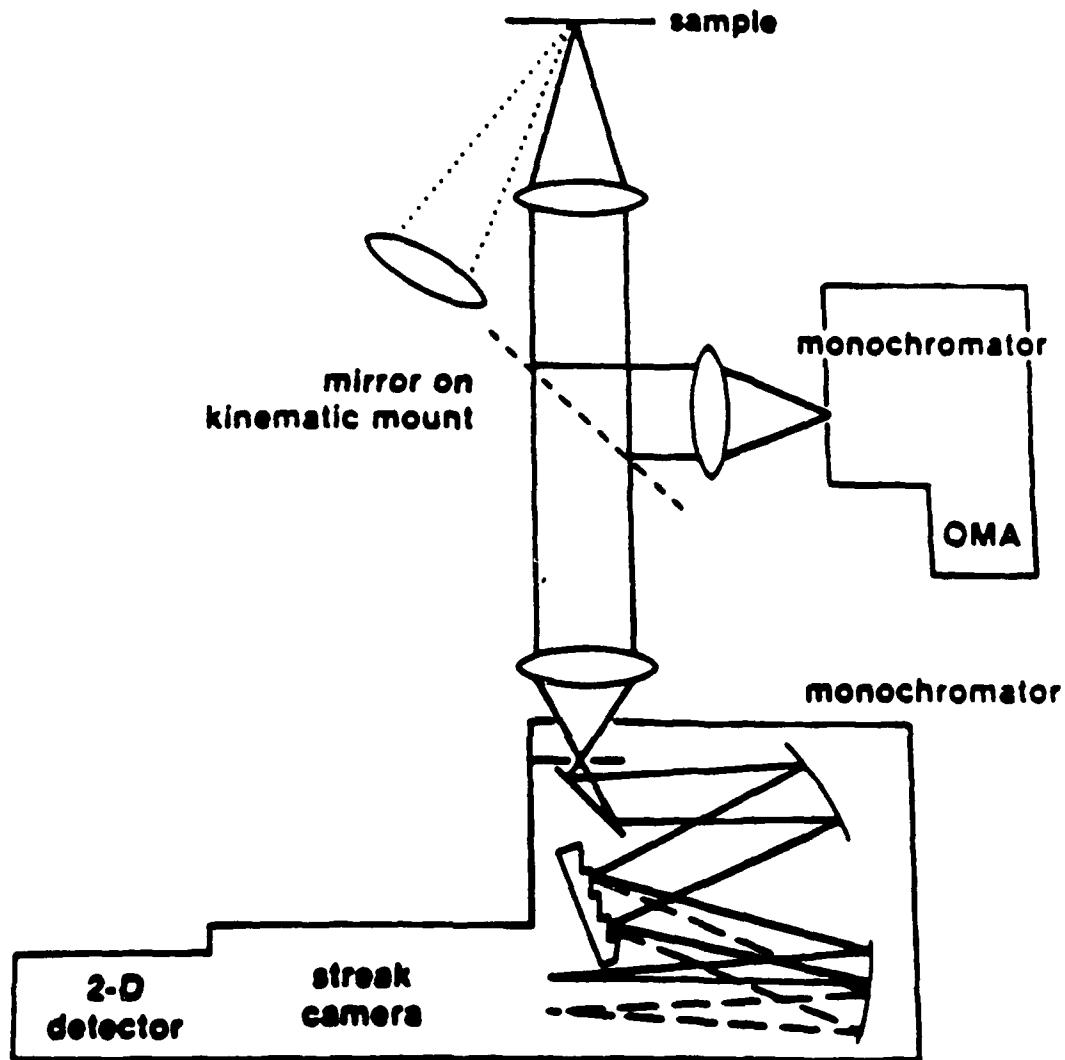


Figure 2: Photoluminescence spectroscopy system. The dye laser pump beam is focused on the sample as shown by the dotted line. The dashed line indicates a mirror on a kinematic mount. The optical beam path through the streak camera monochromator is detailed to show how the luminescence is dispersed across the entrance slit of the streak camera.

### I.A.3 Cryogenics

The PL experiments were always carried out with the sample held at 6K. This was required for two reasons: (i) low temperatures were required for the PL signal level to be strong enough to be detectable by the streak camera detection system with reasonable signal integration times, and (ii), we wanted to avoid as much as possible complications that arise at higher temperatures, such as phonon effects and the effects due to broad carrier distribution functions.

The cryostat was a TRI Research model RC110 flow cryostat. The sample was held in vacuum on an oxygen-free copper cold finger. Electrical vacuum feedthroughs were provided so that a bias voltage could be applied to the sample. A calibrated silicon diode (Cryocal model DT-500) was used to monitor the temperature of the cold finger near the sample; for all the experiments reported here the temperature was 6K.

The laser beam was focused onto the sample surface, usually with a 152-mm-focal-length lens. The spot size of the pump beam at focus was measured by scanning a 12.5- $\mu$ m pinhole across the beam at the same position as the sample. The pinhole was scanned using a computer-controlled stepper with 1- $\mu$ m resolution. The light passing through the pinhole was detected with a PIN diode, integrated on an A/D converter, and stored in the computer. (A description of the data acquisition software may be found in ref. 6). The spot diameter was typically 30-60  $\mu$ m, though the error in the measurement was about -5%, +40%, due primarily to the difficulty of positioning the pinhole at precisely the location of the sample in the cryostat.

### I.B. Subpicosecond Absorption Spectroscopy

In this section we describe in detail the laser system we have developed in our laboratory over the last several years that is now capable of performing time-resolved absorption (or reflection) spectroscopy over the entire visible spectrum with 100-fs resolution, and with kHz data acquisition rates. Such experimental capabilities have been developed only in the last few years; a good recent review of the state-of-the-art techniques of amplified femtosecond lasers has been provided by Knox.<sup>3</sup> We will first discuss the laser system components used to generate a white-

light continuum-pulse, and then describe in some detail the techniques used to perform pump-probe experiments using these short continuum pulses.

### I.B.1. Dye Oscillator

The first requirement of a short-pulse white-light laser system is, of course, an oscillator capable of producing pulses in the 100-fs range or shorter. Such pulses can be achieved using pulse-compression techniques to generate femtosecond pulses from picosecond dye oscillators.<sup>4</sup> However, aside from the fact that such techniques had not been well-developed at the time we began work on the dye oscillator, the stability requirements of the short pulse source for the generation of a stable, useful, white-light continuum are extremely stringent, and the usefulness of this technique to generate a stable continuum has yet to be demonstrated. Hence the requirement for a femtosecond dye oscillator.

The most widely used source of femtosecond optical pulses is the colliding-pulse-mode-locked dye laser (CPM)<sup>5</sup>, which typically produces pulses in the 60-100 fs range, although pulses as short as 27 fs have been achieved.<sup>6</sup> In its usual ring configuration, the CPM is pumped by a cw argon-ion laser. The ring contains separate gain and saturable absorber jets. The laser operates in its minimum-loss condition, in which there are two pulses counterpropagating in the ring which collide in the saturable absorber. The pulses interfere in the absorber jet, so the absorber is saturated more deeply. This serves to shorten the pulses and stabilize the laser output. The pulses can be amplified by pumping a dye amplifier with nanosecond pulses from a frequency-doubled Q-switched Nd:YAG laser<sup>7</sup> or from a copper-vapor laser.<sup>8</sup>

In our laboratory, however, we have taken an alternate approach to amplification of ultrashort dye laser pulses, in which we pump the dye amplifier with short (<100 ps) pulses.<sup>9,10</sup> Pumping with short pulses results in an efficient high-gain amplifier with good contrast between the amplified pulse and amplified spontaneous emission, and also makes the further amplification of the white-light continuum straightforward. This scheme of amplification requires that the dye oscillator and amplifier pump pulses be synchronized to within a few tens of picoseconds, hence

we term the "synchronous" amplification. Synchronous amplification requires that the dye oscillator be synchronously pumped. The dye oscillator we describe here is synchronously pumped, but has the additional advantage of being colliding-pulse mode-locked. We have further discussed the design and development of this laser in references 10 and 11.

A necessary condition for a dye laser to be both synchronously pumped and colliding-pulse mode-locked is that the critical position of the saturable absorber be unaffected by adjustments of the cavity length. The use of an antiresonant ring as one end mirror of a linear cavity<sup>12</sup> enables this condition to be satisfied. An antiresonant ring consists of a 50% splitter and a ring to return the beams to the splitter. An incoming pulse from the linear part of the cavity is split into two equal parts by the 50% splitter; when the two pulses recombine on the splitter, they interfere so that the entire pulse is returned to the cavity. In fact, as Siegman has shown,<sup>12</sup> if the splitter reflectivity is  $R$  and the transmission  $T$ , then the power returned to the cavity is given by  $4RT$ , and the power coupled out of the cavity is  $(T-R)^2$ . As suggested by Siegman<sup>13</sup> and subsequently demonstrated with Q-switched Nd:YAG<sup>14</sup> and Nd:Glass<sup>15</sup> lasers, colliding-pulse mode-locking may be obtained by situating a saturable absorber exactly opposite the 50% splitter in the antiresonant ring. Our laser represents an extension of this technique to the synchronously-pumped, cw mode-locked regime.

A diagram of the dye laser is shown in Fig. 3. The laser consists of a four-mirror linear cavity, for which one end mirror is a 5% output coupler mounted on a translation stage and the other is an antiresonant ring. The gain medium is a 200- $\mu\text{m}$  jet of Rhodamine 6G in ethylene glycol, and the saturable absorber is a 20- $\mu\text{m}$  jet of DODCI (Diethyloxadicarbocyanine Iodide) in ethylene glycol. The typical DODCI concentration is  $2 \times 10^{-3}$  M, which causes the laser to operate with a central wavelength of 615-618 nm.

The dye laser is synchronously pumped at 100 MHz by the frequency-doubled output of the cw mode-locked Nd:YAG laser. Both the Nd:YAG pump laser and the dye laser are mounted on Super-Invar slabs to minimize cavity-length fluctuations. The shortest pulses and greatest stability are obtained when the cavity lengths are properly matched and the Nd:YAG laser

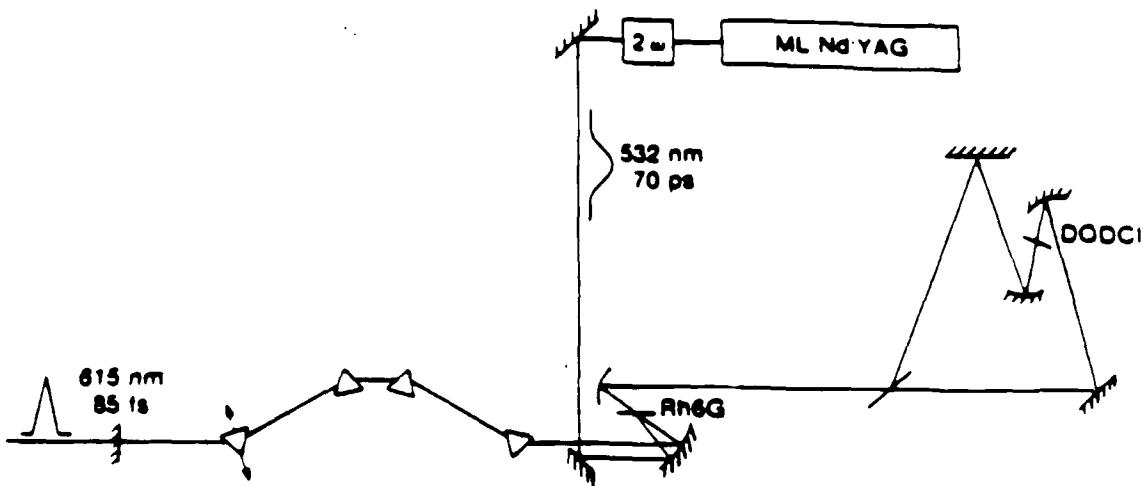


Figure 3: Synchronously pumped, colliding-pulse mode-locked antiresonant ring dye laser. The arrows on the first prism indicate how the prism is moved to tune the intracavity dispersion.

cavity length is adjusted so the phase jitter of the Nd:YAG output is minimized. Both the pump and dye lasers also have end-mirrors mounted on piezo-electric transducers to make the required submicron cavity length adjustments easy and reproducible. When the cavity lengths are well matched, the dye laser stability is of the order of the pump laser stability, which for frequency-doubled Nd:YAG is about 1% rms.

The dye cavity mirrors have single-stack high-index coatings centered at 620 nm to minimize the effects of unwanted temporal dispersion.<sup>16</sup> In order to control the dispersion in the cavity, four quartz Brewster prisms are situated in the linear part of the cavity. As first demonstrated by Fork *et al.*<sup>6,17</sup> in a cw-pumped CPM laser, the angular dispersion of the prisms introduces negative group velocity dispersion (GVD), while the prism glass introduces positive GVD. Therefore, by controlling the amount of glass in the cavity (*i.e.*, by moving one of the prisms in or out of the cavity along its axis), one may tune the net GVD of the cavity. The dependence of the pulse width on the intracavity glass is shown in Fig. 4. It should be noted that not only is there a sharp minimum in the pulse width but also the pulse shape and spectrum differ

qualitatively in two different regions of the graph. When there is too little glass in the cavity, the pulse has negligible wings, the spectrum tails to the yellow, and the time-bandwidth product  $\tau_p \Delta v$  is 0.35. When there is too much glass in the cavity, the pulse has broad wings, the spectrum tails to the red, and  $\tau_p \Delta v$  is 0.5. The minimum pulse width occurs with approximately 50-100  $\mu\text{m}$  less glass than the prism position which yields a symmetrical spectrum. This is consistent with a picture of pulse shaping where the minimum pulse width occurs when positive self-phase modulation (SPM) is balanced by negative GVD from a stable pulse. The dependence of the laser spectrum on the intracavity glass is shown in Fig. 5. Generally the laser is most stable when operated with a slight net negative cavity GVD (*i.e.* with slightly less glass than that which gives

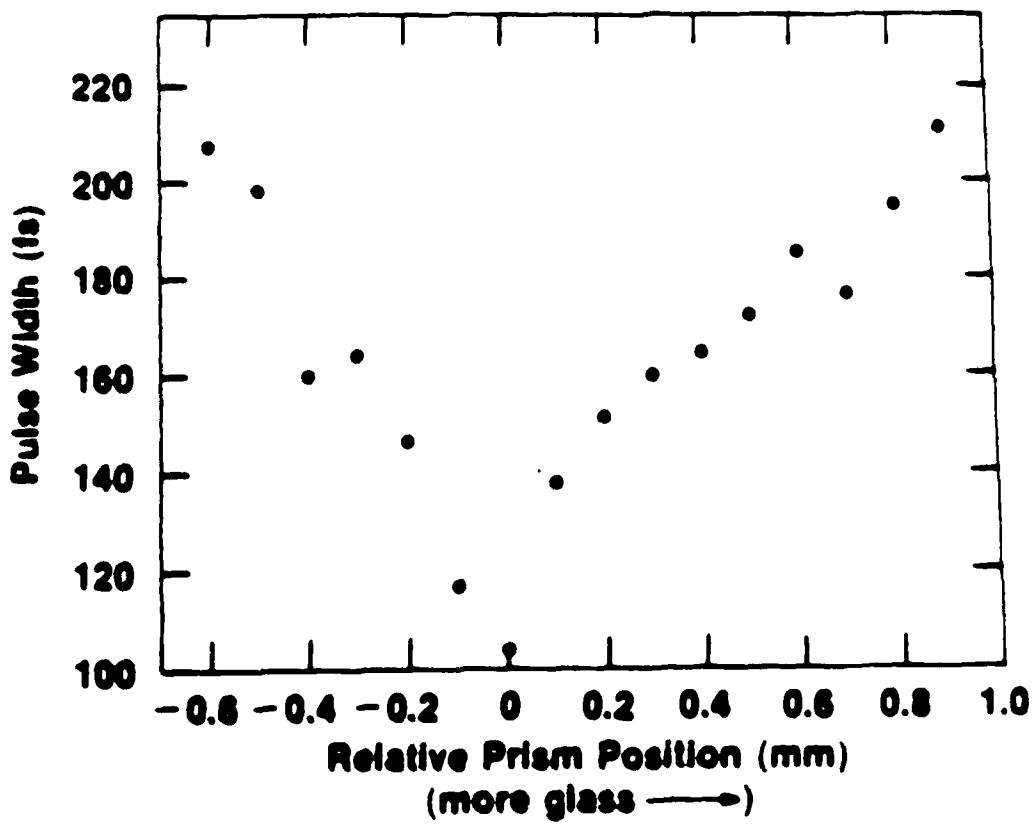


Figure 4: Pulse width of the antiresonant ring laser vs intracavity glass.

the shortest optical pulse). Typically the laser pulse width is about 90 fs (assuming a  $\text{sech}^2$  pulse shape), determined by background-free autocorrelation, but pulse widths as short as 45 fs have been obtained.

### I.B.2 Nd:YAG Regenerative Amplifier

The pump source for the dye laser amplifier is the frequency-doubled output of a Nd:YAG regenerative amplifier. The regenerative amplifier is cw-pumped, so that the laser can be Q-switched, injected with a short (80 ps) seed pulse, and cavity-dumped at a 1 KHz repetition rate. A schematic drawing of the amplifier configuration is shown in Fig. 6. The basic idea of the synchronous dye amplifier is to see the Nd:YAG regenerative amplifier with a pulse from the cw

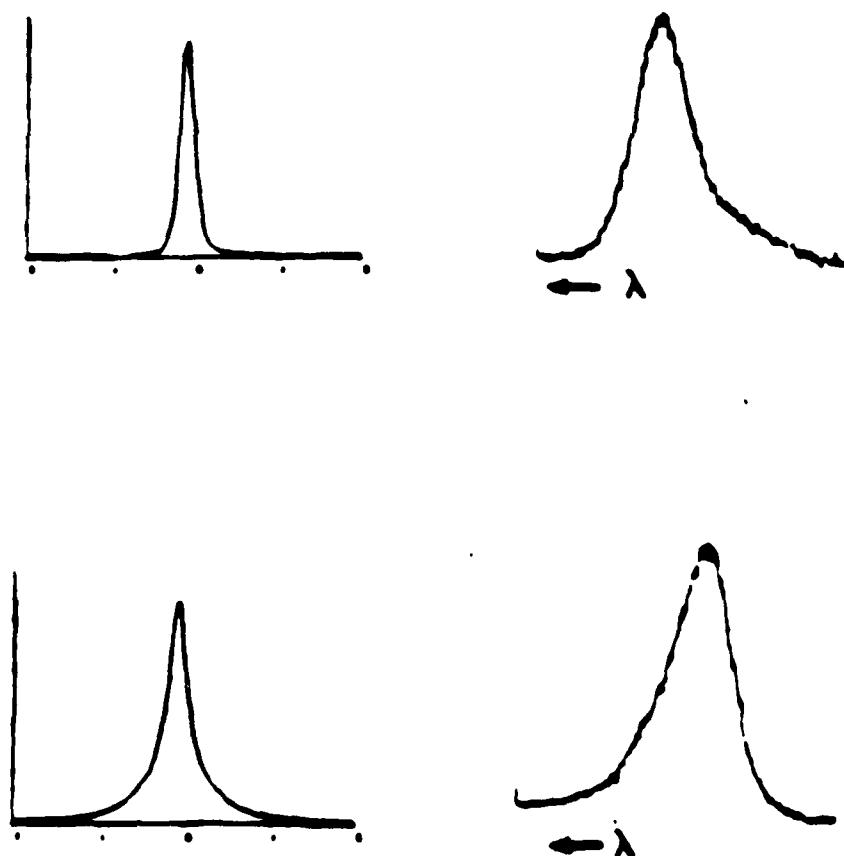


Figure 5: Pulse autocorrelations and spectra with two different laser conditions.  
Top: net negative intracavity GVD. Bottom: net positive intracavity GVD.

mode-locked Nd:YAG that is the pump source for the dye oscillator. This guarantees that the amplifier pump pulse and the dye laser pulse will be strictly synchronized. Timing between the pump and signal pulses is then simply accomplished with an optical delay line.

A portion of the 1.06- $\mu$ m radiation from the Nd:YAG oscillator is injected into an optical fiber. The fiber is used simply to transport the beam across the laboratory to the amplifier and to decouple the alignment of the Nd:YAG oscillator and regenerative amplifier. After the fiber, a Pockels cell switchout selects a single pulse from the 100-MHz train of pulses. This pulse is injected into the regenerative amplifier off a 4% reflector. The injected pulses are of approximately 100-pJ energy and 80-ps duration.

The regenerative amplifier design is shown in Fig. 6. The laser head is a Quantronix model 117 with 3 x 104 mm Nd:YAG rod. The thermal lensing of this rod is very strong, and a cavity round-trip time of about 12 ns is desirable so that any secondary pulses in the cavity due to injection of unwanted pulses from the oscillator will be separated in time from the main pulse. Furthermore, ideally the beam in the intracavity Pockels cell should be large enough to prevent damage to the Pockels cell crystal, and be collimated to enhance the switchout contrast ratio. The contrast ratio can be further enhanced by using the thin-film polarizers in reflection mode. The laser cavity shown in Fig. 6 satisfies all these requirements.

The operation of the regenerative amplifier is as follows. The quarter-wave plate in the cavity frustrates the cw lasing and rejects any injected light that leaks through the first switchout after two round trips. The amplifier is triggered by applying a voltage step to the lithium niobate Pockels cell sufficient to compensate for the quarter-wave plate. This action not only Q-switches the laser but also traps the injected pulse (selected by the first switchout) in the cavity. Following approximately 40 round trips (480 ns), a second voltage step is applied to the Pockels cell providing an additional quarter-wave rotation, and the pulse is rejected from the cavity.

The result is that the regenerative amplifier output is routinely 1.2 mJ at a repetition rate of up to 1.7 kHz, with a pulse width of 80 ps. The repetition rate is limited entirely by the first

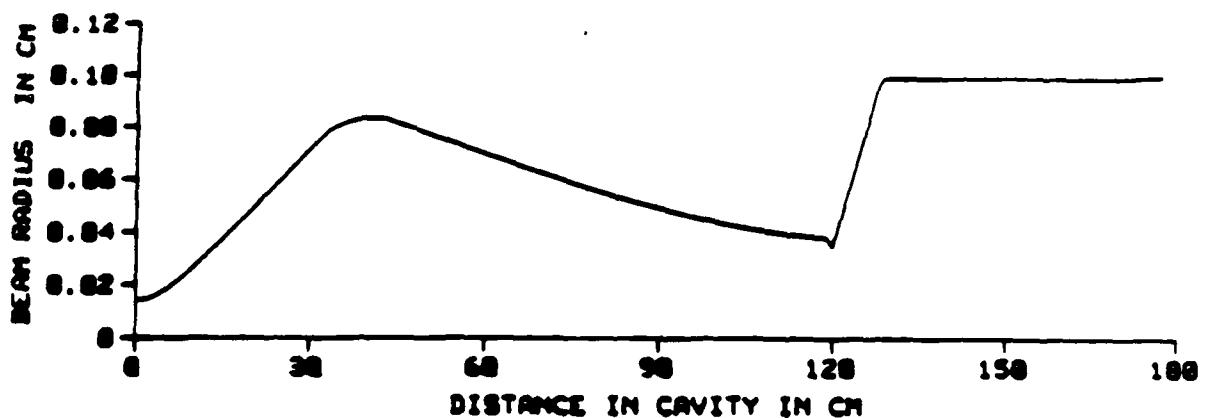
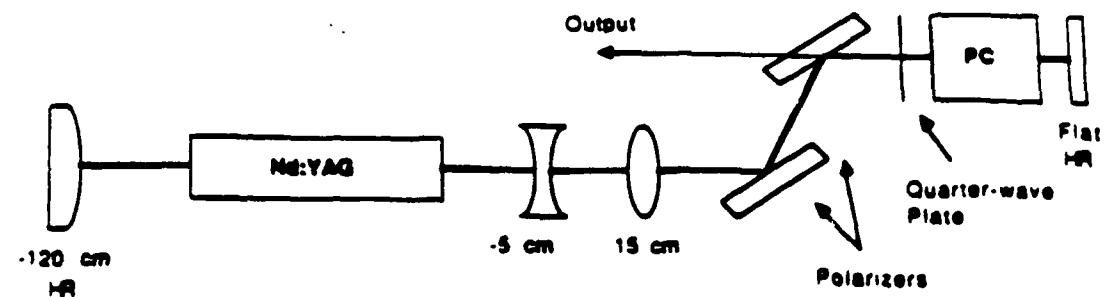


Figure 6: Nd:YAG cw-pumped regenerative amplifier. On top is a schematic of the cavity design. Below is the calculated beam diameter in the cavity.

switchout; typically the laser is operated at 1 kHz. The pulse energy is limited by the thermal depolarization in the Nd:YAG rod combined with the use of high-contrast dielectric polarizers in the cavity. Output energy fluctuations are typically less than 2% rms.

The output of the regenerative amplifier is down-collimated by a telescope to a diameter of approximately 500  $\mu$ m, and is frequency-doubled in a 2.5 cm CD\*A crystal which is temperature-tuned to phase match at 50°C. The second-harmonic energy is 500  $\mu$ J per pulse. The second-harmonic is directed by a dichroic mirror to the dye amplifier, the remaining fundamental is again down-collimated and frequency-doubled in an 8-mm KTP (potassium titanyl phosphate) crystal.

This second-harmonic, which is used to amplify a portion of the white light continuum, has an energy of  $80 \mu\text{J}$  per pulse.

### I.B.3. Dye Amplifier

The dye amplifier is configured as a standard two-stage, collinearly pumped amplifier.<sup>18</sup> The stages are 1 cm in length with flowing dye solution. We found the optimum dye solution to be Sulforhodamine 640 dissolved in a 50/50 mixture of methanol and water. This dye solution offers the best combination of high gain but low amplified spontaneous emission (ASE) level, essentially because the ratio of water to methanol can be varied so that the gain spectrum of the amplifier dye matches the dye oscillator spectrum. (A higher proportion of water, which is a polar solvent, produces a gain spectrum which is shifted to the red). The dye concentration is about  $5 \times 10^{-5} \text{ M}$ .

The spot sizes and relative pump power for each amplifier stage were carefully optimized so that the maximum gain could be extracted from the amplifier consistent with a minimum of pulse broadening due to gain saturation. The optimized amplifier has a first stage pumped by 20% of the second-harmonic energy from the frequency-doubled Nd:YAG regenerative amplifier. The spot size of the pump and oscillator beams is about 150  $\mu\text{m}$  diameter, and the gain is typically 1000. The amplified dye beam then passes through a 200- $\mu\text{m}$  jet of Malachite Green saturable absorber, which serves to absorb the unamplified 100-MHz dye pulses and most of the ASE from the first stage. The amplified beam suffers a loss of about two in the absorber. The second stage is pumped by the remaining second-harmonic. The spot size is about 1-mm diameter, and the gain is typically 100. Thus, including the absorber loss, the net gain is  $5 \times 10^4$ , so since the input pulse energy from the oscillator is about 200 pJ, the amplified pulse energy is 10  $\mu\text{J}$ . (This corresponds to the measured average power of 10 mW). The highest pulse energy we have observed with this system without pulse broadening is 15  $\mu\text{J}$ . If pulse width is not a concern, the spot sizes in the two stages can be reduced so the amplifier will be driven further into saturation. This can produce amplified pulses with energy in excess of 40  $\mu\text{J}$ , but the pulse then broadens to about 200 fs FWHM.

The ratio of the amplified short pulse energy to ASE at the output of the amplifier is typically 1000:1, though this figure varies significantly with the age of the amplifier dye and the relative water to methanol content of the solvent. This figure is of concern only when a pump-probe experiment is being performed where a pump pulse near 615 nm is necessary, in which case the effect of the ASE portion of the pump pulse must be considered. In all the experiments reported here, the amplified dye pulse was used to generate a white light continuum, from which a pump pulse far to the red of 615 nm was selected, so the ASE from the amplifier was not an important consideration.

Aside from avoiding saturation of the amplifier to maintain a short dye laser pulse, it is necessary to compensate for the GVD of the amplifier chain. This is accomplished by passing the oscillator output through a flint (SF-10) prism pair in a double-pass configuration before the amplifier chain. The pulse width at the output of the amplifier is then easily minimized by adjusting the amount of glass the beam travels through in the prism pair (in exactly the same fashion as in the dye oscillator). Thus the prism pair puts a slight negative chirp on the dye pulse that is exactly compensated by the positive GVD of the amplifier chain, and the output of the amplifier is typically 90-120 fs. The prism pair is placed before the amplifier so that the loss due to reflections and scattering in the prisms can be recovered in the amplifier. (If it were placed after the amplifier, there would be less energy available for the continuum generation). As a further precaution against GVD that cannot be compensated with the prism pair, the mirrors in the amplifier chain all have single-stack dielectric coatings centered at 615 nm.

The stability of the amplified dye pulse is 3% rms (measured by taking a histogram of the pulse energy). It is extremely important for the generation of a stable, useful continuum that the amplified dye stability be less than or of the order of 3%.

#### I.B.4. Continuum Generation and Amplification

The amplified dye laser pulse is focused with a 65-mm lens onto a 1-mm jet of ethylene glycol. A thick optical-quality jet is obtained using a low-pressure flow system with a high-quality

sapphire nozzle (Precomp model PRE10N015). The resultant intensity is about  $10^{13}$  W/cm<sup>2</sup>. This intensity for a 100-fs pulse is sufficient to produce a white light continuum, which extends over the entire visible range.<sup>19,20</sup> Apart from the effect of GVD, this continuum pulse is as short as the pump pulse, namely 100 fs. The spectrum of the continuum peaks strongly at the pump pulse wavelength and falls off roughly exponentially to each side, though under certain conditions the spectrum shows structure due principally to self-phase modulation.<sup>21</sup> The position of the dye jet relative to focus is critically important for obtaining a smooth, unmodulated, stable spectrum. Usually the optimum glycol jet position is just beyond focus. The focal length of the pump lens is also critical; a longer lens will produce a modulated spectrum since the continuum generation will be closer to threshold. A shorter lens will produce an intensity that is too high, and the dye laser beam will actually atomize the front and back surfaces of the glycol jet, and a mist of ethylene glycol will be blown off the surfaces. This scatters the laser beam, destroying the beam quality and introducing instabilities in the continuum spectrum and energy.

A portion of the white light spectrum used in the experiments is shown in Fig. 7. (The spectrum is not corrected for the responsivity of the OMA detector). This spectrum is an integration of 5000 laser shots. When the laser system is running properly, the shot-to-shot fluctuations of any given spectral portion of the continuum are about 8-12% rms.

The white light is collimated by a 38-mm achromatic lens after the glycol jet. A beam splitter reflects about 50% of the light into the pump arm of the pump-probe setup, as is shown in Fig. 8. A filter selects a 10-nm bandwidth portion of the spectrum. This beam is sent through a computer-controlled delay line (1- $\mu$ m resolution), and amplified in a 1-cm dye flow cell. The 80- $\mu$ J second-harmonic from the second (KTP) frequency-doubling of the Nd:YAG regenerative amplifier is used to pump this amplifier. A filter after the amplifier is used to block any remaining pump light (and, when possible, ASE from the amplifier). The amplified continuum energy is typically in the 10-70 nJ range in the spectral region of interest in the experiments (690-820 nm). The ASE is at least 100 x below the amplified pulse energy.

The remaining portion of the white light after the splitter is attenuated and used as a probe beam. Typically the probe energy is below the pump energy by at least 100 x. We discuss how pump-probe experiments are performed in the next section.

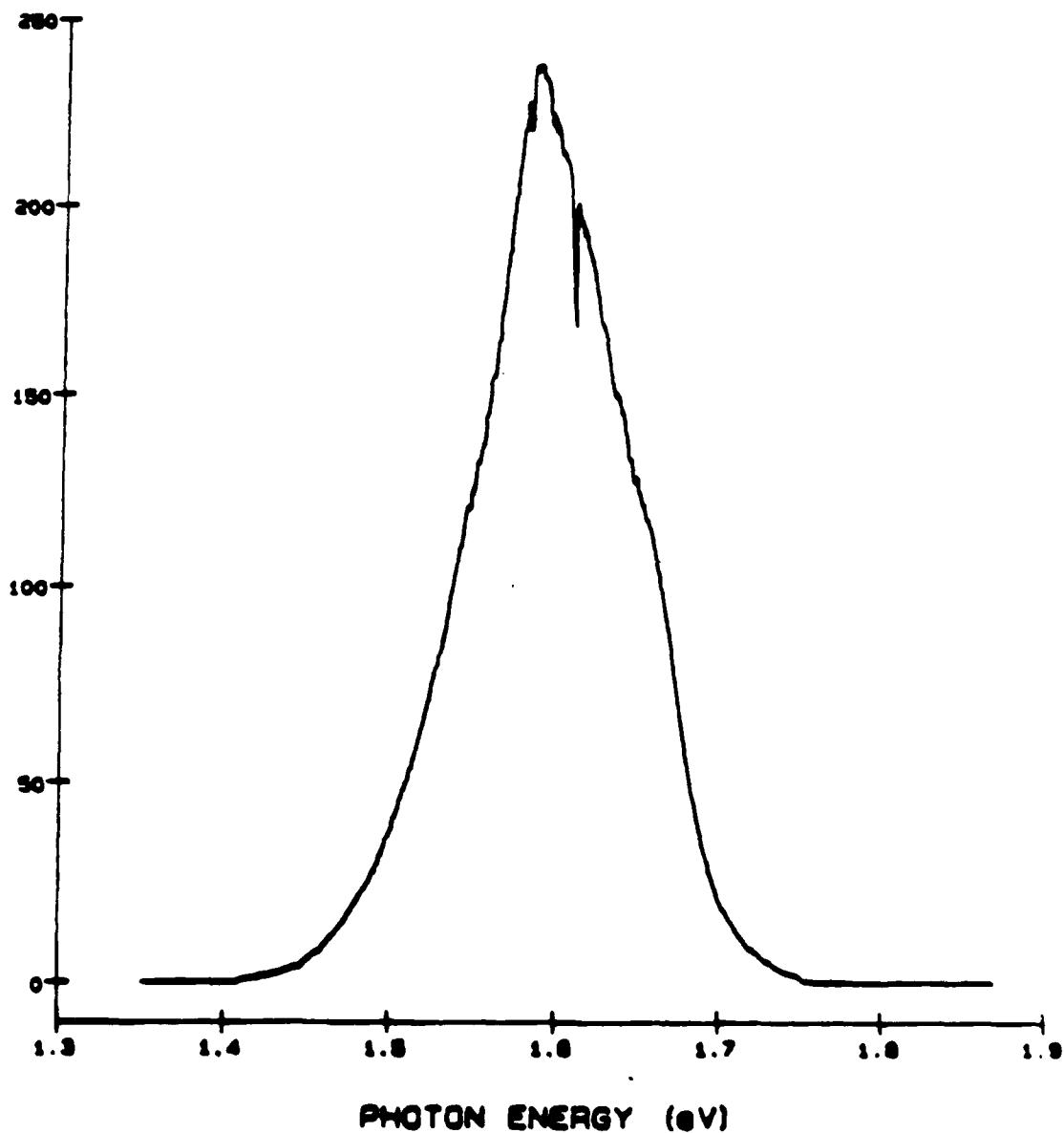


Figure 7: Typical white light spectrum over the probe range 1.4-1.8 eV. An RG780 filter (with transmission edge between 1.6 and 1.7 eV) was used to block the intense portion of the continuum at higher photon energies.

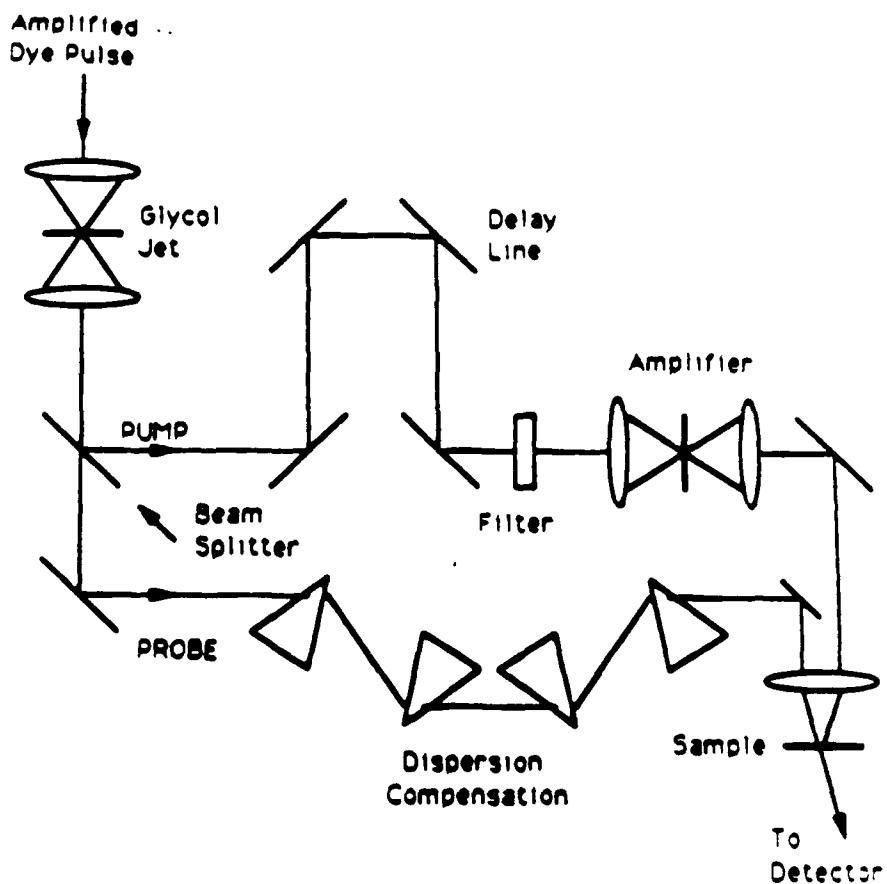


Figure 8: Schematic of the pump-probe setup. Not shown are the pump beam for the continuum amplifier, the neutral density filters in the probe beam, or the probe reference beam detector.

### I.B.5. Pump-Probe Experiments

The basic idea of a pump-probe experiment is very simple. A short optical pulse is used to pump the sample and induce a change in its optical properties, such as its reflectivity or absorption spectrum. A much weaker pulse is delayed in time and used to probe the change in optical properties of the sample due to the pump. The time resolution is obtained by varying the time delay between the pump and probe. Since the pump and probe pulses are derived from the same source by using a beam splitter, there is no temporal jitter between the pump and probe; hence the time resolution is limited only by the optical pulse width.

When a spectral continuum is available for the probe, then the change in the entire spectral region of interest due to the pump can be accumulated at once by using parallel detection: the probe beam after passing through the sample is dispersed through a monochromator and integrated on an OMA. Thus we can do time-resolved spectroscopy. When the change in optical properties due to the pump is small and higher sensitivity is required, the OMA may be replaced with a narrow slit and photomultiplier or photodiode detector. Then the detection is at a single wavelength, but the detectors are fast, so that if the probe beam is simultaneously monitored before it goes through the sample, energy fluctuations may be normalized out on every single laser shot, and a substantial sensitivity increase may be obtained.

The time origin, i.e. the delay line position for which the pump and probe beams arrive at the sample at the same time, is precisely determined by cross-correlating the pump pulse with the probe pulse. A nonlinear crystal ( $\text{LiIO}_3$ ) is placed at the same position as the sample. The pump and probe beams are aligned so they are parallel and about two beam diameters apart. They are focused on the crystal, which is aligned to generate the sum frequency of the pump and probe pulses. The sum frequency beam is dispersed in a monochromator before being detected by a photomultiplier tube. By tuning the monochromator wavelength, one may easily select which spectral portion of the probe is to be correlated with the pump. The sum frequency signal is detected as a function of time delay, the result being the cross-correlation; a typical example is given in Fig. 9. The FWHM is 180 fs. The pump and probe pulse widths are likely very close, so to a good approximation the cross-correlation may be deconvolved to get the real pulse width in the same fashion as an autocorrelation. For  $\text{sech}^2$  pulse shape, the deconvolution factor is 0.6482,<sup>28</sup> so the pulse width is 120 fs.

It is particularly important when doing time-resolved spectroscopy to avoid systematic error due to GVD, since the entire probe spectrum must arrive at the sample at the same time. If the continuum spectral range of interest is very broad, then the temporal shift due to GVD in the glycol jet, lenses, and filters between the red and blue portions of the probe spectrum can be significant. This temporal shift can be directly measured by cross-correlating the probe with the pump over the

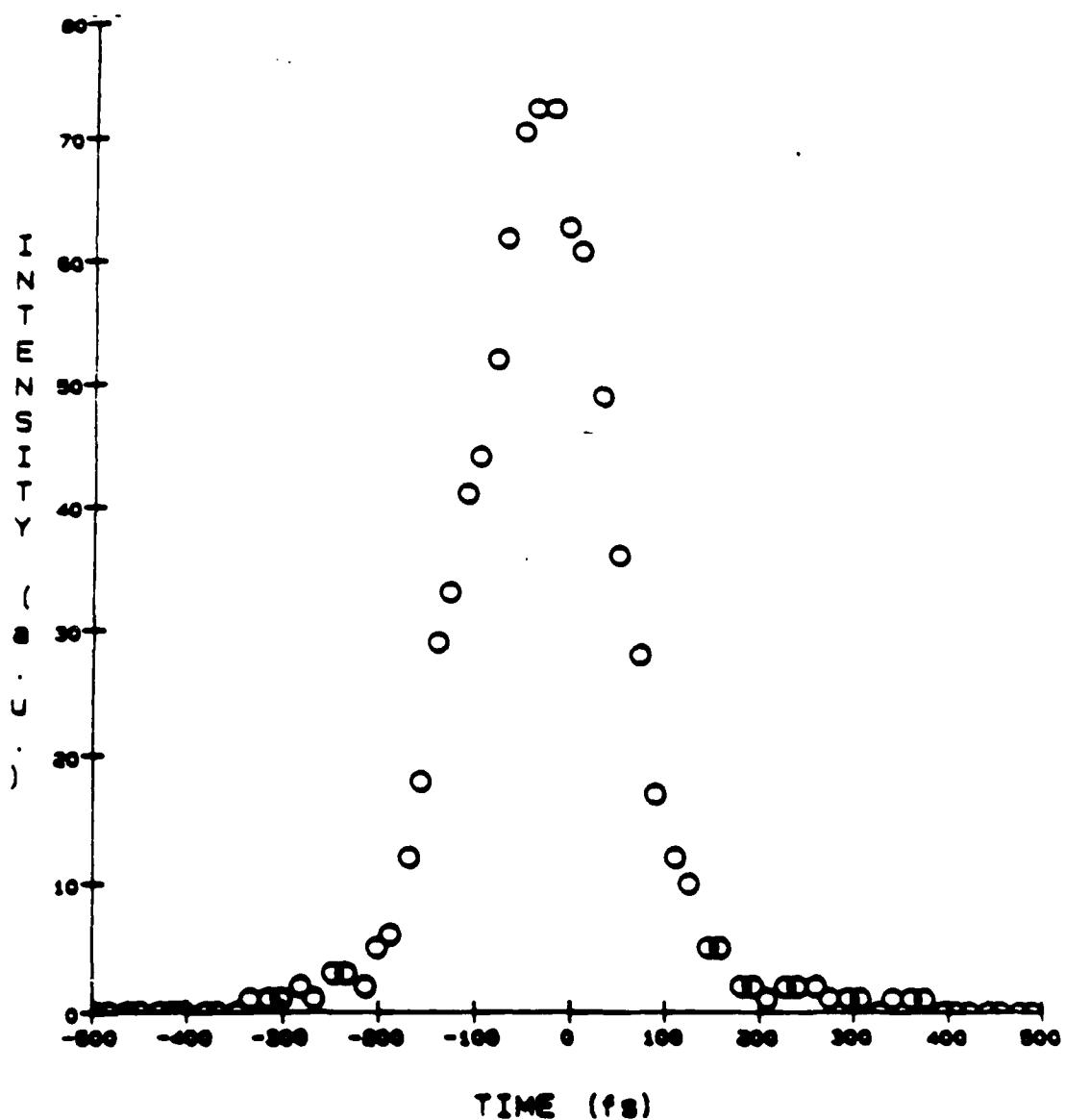


Figure 9: Cross-correlation of the 770 nm portion of the probe with the 790 nm pump beam. The correlation width is 180 fs, which corresponds to a pulse width of 120 fs.

probe spectral range of interest. The result is displayed in Fig. 10 (squares). The pulse width of any given portion (of approximately 10-nm width) of the spectrum is 140 fs, but there is an approximately linear chirp that leads to a 500-ps temporal dispersion over the 120-nm probe spectrum.

This dispersion may be corrected by inserting a prism pair in double pass arrangement in the probe beam. The white light dispersion when quartz prisms are used is shown in Fig. 10 (circles). The pulse width for each spectral component is 120 fs and the dispersion over the 120-nm bandwidth is 30 fs. Unfortunately the quartz prisms are not useful since to obtain the necessary negative chirp, the prisms must be far apart and the beams must go through the prisms as close to the tip as possible. However, in this case the beam suffers from vignetting (due to the several-millimeter beam diameter), and the probe beam quality is destroyed and energy lost. Possibly this problem could be overcome by focusing the beam through the prism arrangement, with the focal position just beyond the second prism. However, a simpler solution is to use flint prisms, which give a much higher dispersion. Thus the necessary prism separation is smaller, and the vignetting problem is eliminated. The continuum dispersion for flint (SF-10) prisms is shown in Fig. 10 (triangles). The pulse width at each spectral component is 110 fs, and the dispersion over the 120-nm bandwidth is 100 fs. This dispersion is somewhat worse than that for quartz prisms, since the flint prisms have a higher quadratic as well as linear dispersion. In fact, to show that the minimum achievable dispersion is limited by the quadratic dispersion of the prisms, we have fitted the dispersion data to parabolas. The result is the dotted curves of Fig. 10. It is important to note that each time the central probe wavelength of interest is changed, it is necessary to adjust the amount of glass the beam passes through in the prism pair to minimize the continuum dispersion.

A further consideration in doing pump-probe experiments is that the probe beam must be centered on the pump beam at the position of the sample. Ideally, the pump beam will be larger than the probe so that the pump intensity is roughly constant over the area of the probe spot. This is accomplished by down-collimating the pump beam after the continuum amplifier. The pump and probe spots are then overlapped by putting a small ( $12.5 \mu\text{m}$ ) pinhole adjacent to the sample. The sample is then placed precisely at the focus by maximizing the throughput of the probe beam through the pinhole; the overlap is then easily obtained by aligning the pump through the pinhole. The overlap may be verified and the spot sizes measured by scanning the pinhole through the

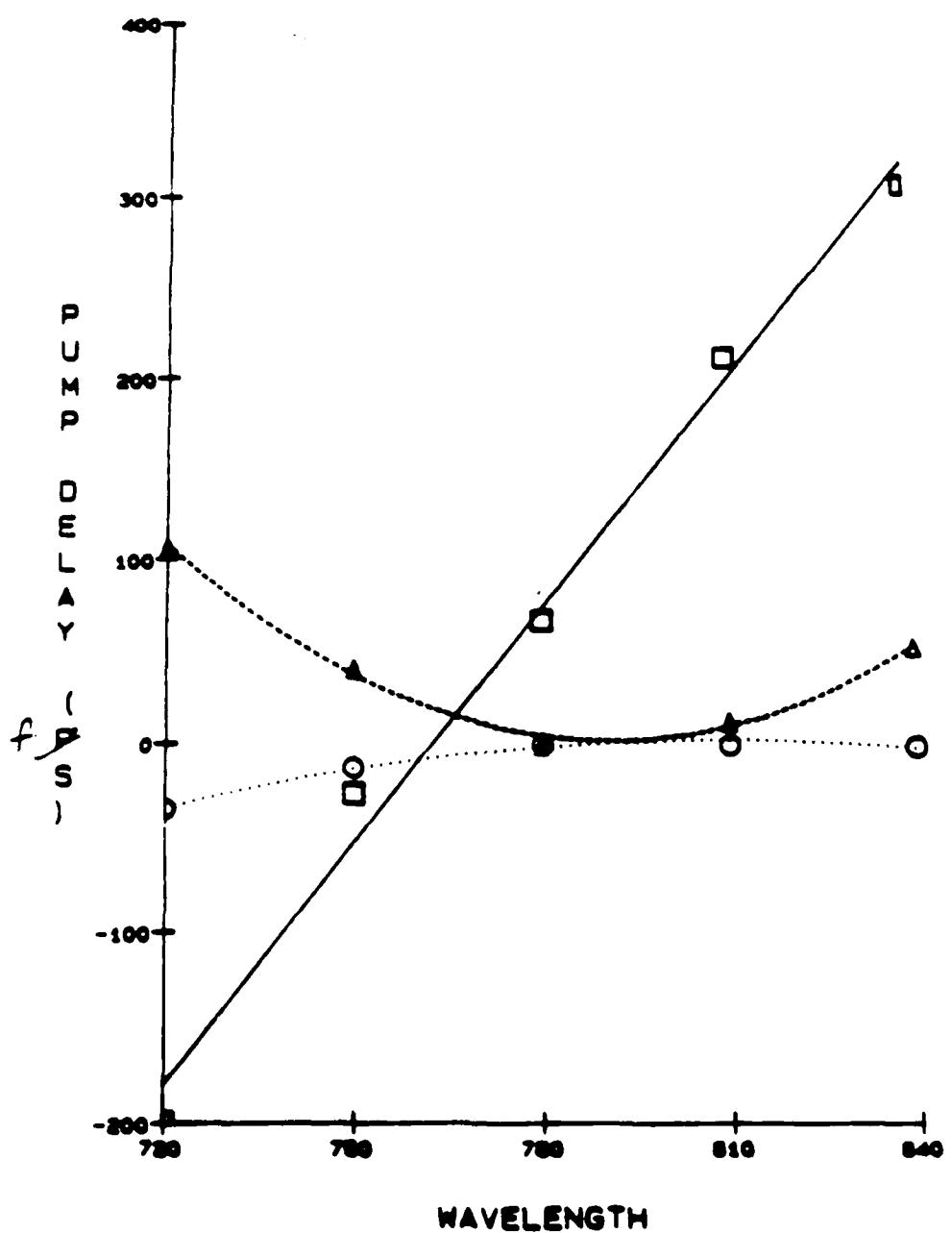


Figure 10: Temporal dispersion of the white-light continuum pulse. The squares show the positive linear dispersion of the uncompensated continuum. The circles (triangles) show the dispersion when quartz (SF-10 flint) prisms are used to compensate for the linear chirp of the continuum.

beams and detecting the transmitted light through the pinhole. For the experiments reported here, the spot sizes were typically a pump diameter of 48  $\mu\text{m}$  and a probe diameter of 26  $\mu\text{m}$ . After the alignment and focus are optimized, the sample is simply translated into place.

#### I.B.6. Data Acquisition

All the white light time-resolved spectroscopy experiments reported in this report measured the so-called differential absorption of the sample. What is measured is  $\Delta T/T$ , where  $T$  is the transmitted probe energy with no pump beam incident on the sample, and  $\Delta T$  is the change in transmission due to the pump, *i.e.*, the transmitted probe with pump incident minus the transmitted probe with no pump incident on the sample.

As we mentioned at the beginning of the previous section, the differential absorption spectrum may be accumulated using an OMA. The readout of the OMA diode array is synchronized to the laser by triggering the readout with a pulse timed to the laser amplifier. The kilohertz trigger from the amplifier timing electronics is divided down typically by 50x, so that the OMA will integrate 50 laser shots and be read out at 20 Hz. Shutters are placed in both pump and probe beams, which may be triggered by the OMA controller. The shutter sequence used to accumulate the differential absorption spectrum is as follows. Both shutters are initially open, and one scan of the transmitted probe is accumulated into memory A. The pump shutter is then closed, and one scan of the probe is subtracted from A. The probe shutter is then closed and a background (dark current) spectrum is added to A. The pump shutter is then opened, and any pump light that is scattered onto the OMA is subtracted from A. (For very high quality samples and large signal levels the last two steps are not necessary). The transmitted spectrum without pump is then accumulated into memory B. An ignored scan is performed between each accumulated scan since a scan readout does not completely reset the OMA diode array. The cycle is repeated as often as necessary to achieve the desired signal-to-noise ratio. Once the data is acquired, it is transferred to the PDP-11 computer and memory A is normalized to memory B; the result is the  $\Delta T/T$  spectrum.

For reasonable (5-30 min) integration times, the typical minimum  $\Delta T/T$  signal observable (i.e., for a single-to-noise ratio of one) is about 0.5%. This is limited by continuum fluctuations and the cavity-length stability of the dye oscillator. It may be possible to improve the sensitivity substantially by stabilizing the dye oscillator cavity length with a feedback loop.

For experiments where it is not necessary to accumulate the entire differential absorption spectrum at once, photodiodes or photomultipliers may be used to monitor the probe light. For these experiments, a portion of the probe is split off before the sample and used as a reference beam. Both transmitted probe ("A") and reference ("B") beams are passed through monochromators to select the wavelength of interest, and detected with photodetectors. For the experiments reported here, the probe energies were so low that photomultipliers with extended red multi-alkali photocathodes (Hamamatsu type R2066) were required. For each laser shot, A and B are integrated and digitized using a LeCroy 2249W pulsed A/D. The computer acquires A and B, checks to see that B is within an acceptable energy range, and computes the signal  $S = (A - B)/B$ . This yields the differential absorption of the sample up to a factor due to the chosen signal level of B. The signal S may be calibrated to yield the quantitative value of  $\Delta T/T$ .

The attainable sensitivity for these single-wavelength experiments is limited by two things. First, the sample surface must be of high quality and the probe beam directionality must be stable. Otherwise the detectors A and B will not observe exactly the same signal and random fluctuations will be introduced. (It is also extremely important that the two detectors be monitoring exactly the same wavelength portion of the spectrum). Secondly, in order to avoid saturating the photocathode output, it is necessary to limit the incident photon flux. I have found it necessary to limit the flux such that only about 625 photo-electrons are emitted from each photocathode for each laser shot. Thus the shot noise fluctuations will be about 4%. For a reasonable experiment, about 20,000 laser shots may be accumulated at each data point, so the shot noise limit to the sensitivity is about  $3 \times 10^{-4}$ . In practice the sensitivity limit is about 0.1%.

### I.C. Description of the Electro-Optic Sampling Technique

The electro-optic sampling (EOS) technique has been chosen as the tool for measuring photoconductively generated transient voltages for a number of reasons. The primary reason is that of temporal resolution; this technique is capable of 100-fs resolution if lithium tantalate is used, and could achieve even higher speeds if other electro-optic materials were utilized. Another reason is the flexibility and noninvasiveness of the technique. A number of embodiments have been developed, including ones that require little or no special modification of the device under test and which can be set up relatively easily. The technique is insensitive to wavelength over a wide band giving the experimentalist complete freedom to choose the most appropriate laser for the research at hand. The development of the technique and its different embodiments will be discussed below.

The EOS concept is based on the linear electro-optic effect. In certain "electro-optic" materials an electric field applied across the material will induce a birefringence. Polarized light passing through the material will be rotated to align itself with the molecules and the applied field. Hence a measurement of the degree of rotation of polarization of the probe light will yield a measurement of the applied field. Specifically, the phase retardance  $\delta$  introduced by a voltage  $\Delta V$  is given by

$$\delta = \pi \frac{\Delta V}{V_\pi} \quad (1)$$

where  $V_\pi$  is the "half-wave voltage" of the material determined by its crystalline structure. If an analyzer, placed after the electro-optic medium, is used to measure the degree of polarization in a particular direction then the intensity of the transmitted light will be

$$I = I_0 \sin^2 \left( \frac{\delta}{2} \right) \quad (2)$$

This relationship is shown in Fig. 11. If the medium is optically or electrically biased at the half-wave point, then a small change in electric field will result in a linear change in transmission.

Hence, after a dc calibration, measuring the intensity modulation of the probe beam will yield a direct measurement of the unknown electric field.

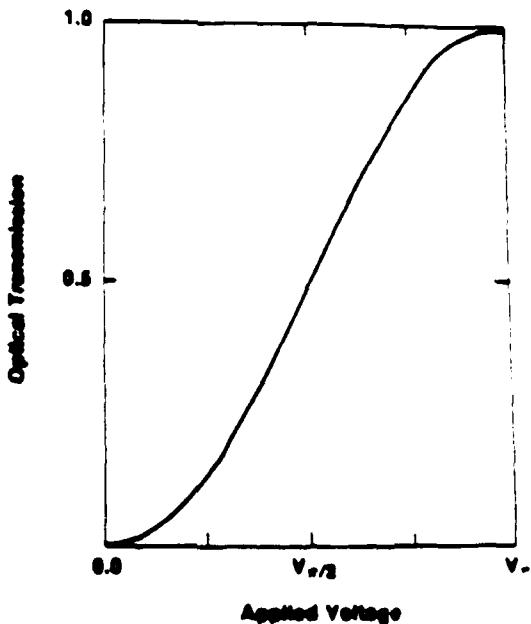


Figure 11: Transmission curve for an electro-optic modulator.

The original EOS test geometry was developed by Valdmanis *et al* 22. In this case the GaAs photoconductive switch is the device under test. In some manner the unknown transient electric field must be coupled into the electro-optic medium. In this first experiment this was accomplished by fabricating a transmission line on lithium tantalate and electrically connecting the photoconductive switch to the transmission line. One laser pulse was used to "close" the photoconductive switch, launching a transient voltage onto the transmission line which then propagated away from the switch. A second synchronized laser pulse was focused into the transmission line and sampled a narrow slice of the transient waveform. The pump and probe pulses are actually pulse trains that are closely synchronized, so that by changing the pump/probe

delay and measuring the modulation of the probe beam as a function of delay the unknown transient voltage as a function of time is obtained.

Valdmanis pioneered the sampling technique and demonstrated useful applications in discrete device testing, characterization of pulse propagation on transmission lines, and measurement of high-speed connector bandwidths 23. He also showed that under most experimental conditions the temporal resolution is determined by the particular sampling geometry and not by the intrinsic response of the electro-optic medium. Specifically, the temporal resolution  $\tau_{\text{exp}}$  is the sum-of-squares convolution of four factors:

$$\tau_{\text{exp}} = [2 \tau_1^2 + \tau_0^2 + \tau_e^2 + \tau_i^2]^{\frac{1}{2}} \quad (3)$$

where  $\tau_1$  is the pump/probe laser pulsewidth,  $\tau_0$  is the transit time of the optical probe pulse across the electric field lines,  $\tau_e$  is the electrical transit time across the probe beamwaist, and  $\tau_i$  is the intrinsic electro-optic response.  $\tau_0$  is determined by the transverse width and the separation of the transmission line electrodes, and  $\tau_e$  is determined by the optical beam spot size. Valdmanis *et al.* showed that by reducing the thickness of the sampling microstrip from 500  $\mu\text{m}$  to 100  $\mu\text{m}$  and reducing the probe beam waist from 110  $\mu\text{m}$  to 15  $\mu\text{m}$  the temporal resolution could be dramatically improved from 2.4 ps to 0.85 ps.

$\tau_0$  is the dominating factor in the microstrip geometry. Because it is not feasible to fabricate microstrips thinner than 100  $\mu\text{m}$  it is necessary to use a different geometry to further improve the temporal resolution. Mourou and Meyer 24 repeated the experiment using a coplanar strip transmission line, shown in Fig. 12. In this case,  $\tau_0$  is governed by the penetration depth of the field into the substrate, which scales with the separation between lines. Using a 50- $\mu\text{m}$ -strip width and spacing, they observed a 0.46-ps rise time. Further improvements in the temporal resolution may be realized by reducing the probe beam spot size to 5  $\mu\text{m}$ , reducing the stripline dimensions to

10  $\mu\text{m}$ , and reducing the laser pulsewidth to 50 fs. Using these parameters and assuming that the intrinsic response of lithium tantalate is 50 fs<sup>25</sup>, Eq. (3) yields a practical resolution limit of 151 fs. All of these refinements are necessary in order to resolve velocity overshoot phenomena on a time scale of a few hundred femtoseconds.

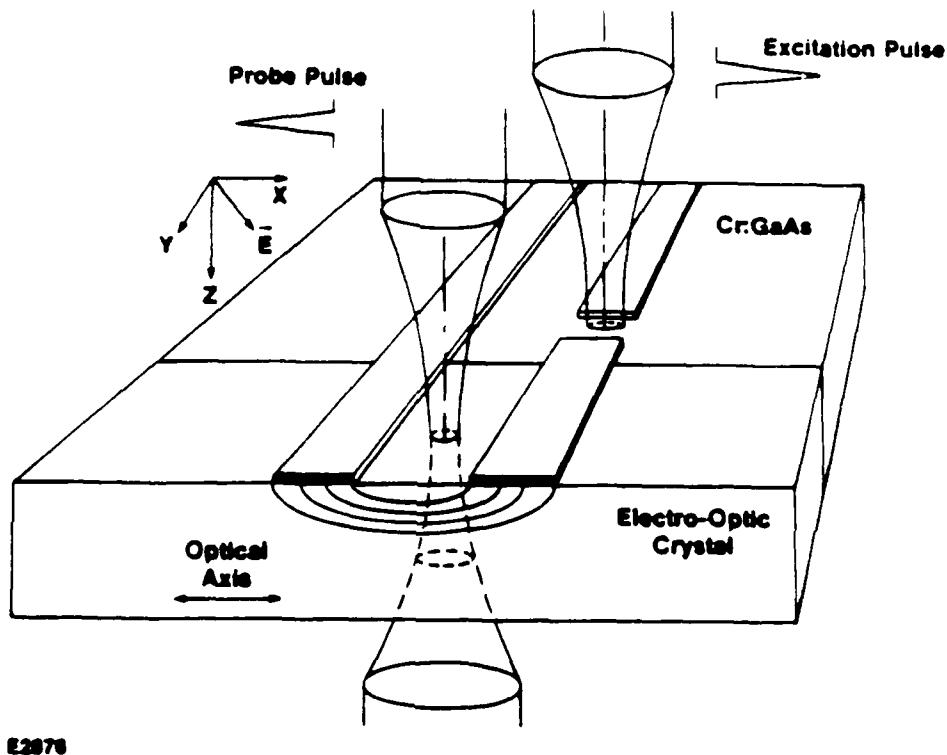


Figure 12: Coplanar stripline sampling geometry<sup>24</sup>.

The EOS technique only requires that some field lines from the device under test extend into the electro-optic medium. It is not strictly necessary to fabricate a separate electro-optic transmission line and then have to worry about efficiently coupling the electrical signal into it (a non-trivial problem). Figure 13 demonstrates several EOS embodiments that have been developed. "In-situ" EOS, (c) in the figure, may be used if the substrate of the device under test is itself electro-optic. GaAs, InP, and CdS are examples of technologically important semiconductor materials that are also electro-optic (Si is a notable exception). Weingarten, *et. al.*<sup>26</sup> and Jai,

et al.<sup>27</sup> have shown that this approach may be used to test internal nodes on GaAs IC's. The drawback to this approach is that the probe wavelength must be sub-bandgap, which for GaAs requires near-IR light. Though picosecond and subpicosecond near-IR lasers have been developed, a stable source of sub-100-fs near-IR laser pulses free of temporal wings has yet to be demonstrated.

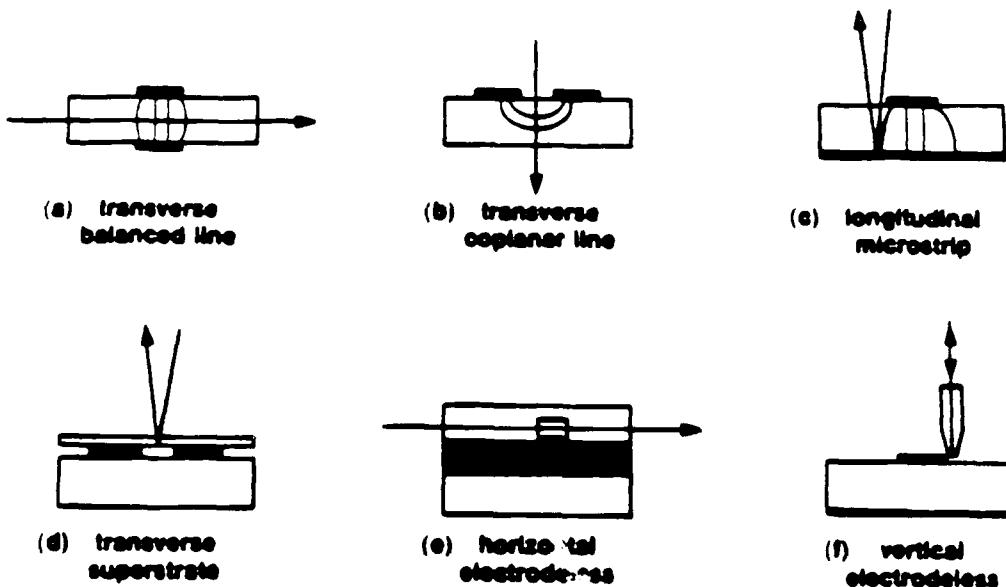


Figure 13: Embodiments of the electro-optic sampling technique.

The electro-optic superstrate geometry, Fig. 13(d), was first described by Meyer and Mourou<sup>28</sup> as a way of performing two-dimensional mapping of transient electrical fields on the surface of a circuit or sample. A high-reflectivity dielectric coating is required in this case on the surface of the EO medium in proximity to the sample. Fringing fields from the sample substrate extend into the superstrate. The probe beam is focused into the area of interest, passes through the fringing field, is reflected off of the HR coating, passes again through the field and exits the sampling crystal. Two-dimensional mapping is achieved simply by scanning the location of the probe beam to different x and y positions of interest. This is the primary geometry used in this

work and will be further detailed in the following section. One disadvantage to this geometry is due to the high dielectric constant of lithium tantalate (40), which may tend to load certain circuits under test. To minimize this loading, the size of the superstrate may be reduced, resulting in the so-called "finger-probe" geometries shown in Fig. 13 (d) and (e). Valdmanis 95 and Nees and Mourou 96 have used these geometries to measure transient waveforms on GaAs- and Si-integrated circuits.

## II. Velocity Overshoot in GaAs

### II.A. Description of the Two Approaches: Time-Resolved Photoconductivity and Transient Absorption Spectroscopy.

The purpose of this work has been to develop and utilize experimental techniques capable of time resolving the dynamics of electrons in GaAs under the influence of an applied electric field during the first few hundred femtoseconds of transport. The specific phenomenon of interest was velocity overshoot, which is of fundamental importance to the operation of a wide range of modern submicron semiconductor devices. Existing experimental evidence for this phenomenon has been reviewed and found to be, at best, incomplete.

The approach taken here has been to concentrate on measurements that may yield direct information about the transient electron drift velocities and energies and that maintain the necessary temporal resolution. The only techniques currently available with sufficient temporal resolution are all-optical or hybrid electro-optic measurements that utilize ultrafast dye lasers. One classic experiment is the measurement of transient photoconductivity, which probes the drift velocity of an ensemble of carriers drifting in a bulk sample by measuring the current induced in the external circuit by the motion of the carriers. This technique has been used extensively from dc down to the nanosecond and picosecond time scales, and in this work has been extended into the subpicosecond regime.

Monte Carlo calculations of the transient electron drift velocity for our particular experimental conditions of temperature, electric field, excitation wavelength, and carrier density

have been developed. They have been compared with the temporal evolution of voltage waveforms generated by a GaAs photoconductive switch. It has been shown in a qualitative analysis that under conditions in which the electron and hole densities and the hole velocities are approximately constant, which is true for this experiment, the time dependence of the voltage transient is determined by the temporal evolution of the electron drift velocities. Hence the measured temporal behavior of the voltage waveform and the calculated time dependence of the electron drift velocity may be compared directly.

Calculations and experiments have been compared with excitation wavelengths of 620 nm and 760 nm and under conditions of high and low applied fields. All of the qualitative features of the Monte Carlo theory have been observed in the experiment. In particular, a significant photocurrent overshoot has been measured at  $E = 10$  kV/cm which is not apparent at lower fields. This constitutes the first fully time-resolved qualitative measurement of velocity overshoot in GaAs. In addition, the Jones-Rees effect, which is apparent as a delay in the onset of photoconductivity at low fields, has also been measured.

Transient absorption spectroscopy is an all-optical technique which measures the shape of the joint electron and hole distribution function. To date this technique has been used to study the relaxation of a hot photoexcited distribution to its thermalized equilibrium state, in the process yielding important information about subpicosecond scattering processes in GaAs. A slight modification of the standard sample structure (addition of an  $n^+$  layer and metallization) allows the experiment to be performed with an applied electric field. Thus heating of the electron distribution by the field may be studied, and this is the process which drives velocity overshoot. Monte Carlo calculations of the evolution of the electron distribution function with various applied fields have been developed and compared with the experimental results. An increased population in the high-energy tail of the distribution has been consistently observed which takes place on a time scale of 100-200 fs. Due to difficulties with the samples which resulted in nonuniform field distributions in the structures, the observed effects were not as large as predicted.

### II.B.1. Predictions of the Monte Carlo Theory

The Monte Carlo model used here is a standard one that has been applied to the study of transport in GaAs. The band structure includes three nonparabolic valleys ( $\Gamma, L, X$ ) in the conduction band and a single parabolic heavy hole band. The light hole band and split-off hole band have been neglected due to the low density of states in those bands. Scattering from elastic acoustic phonons, inelastic optical phonons, and ionized impurities have been included. All simulations for conditions of the transient photoconductivity experiment have been run with  $T = 300K$  and have assumed a density of  $n = 1 \times 10^{17} \text{ cm}^{-3}$ . Osman and Ferry<sup>31</sup> have shown that at this moderate density electron-hole scattering may be neglected. Photogeneration of the electron/hole gas is assumed to be a delta function in time and in energy (determined by the wavelength of photoexcitation). One subtlety of the photogeneration process is that, unlike for carrier injection from contacts, this generation process results in an initial state of the carriers which has no predetermined net momentum. The calculation therefore initializes the carriers with a fixed energy but with a  $k$ -vector chosen at random.

The simulation models transport in the bulk without detailed consideration of what happens when carriers reach the contacts. This is justified if the transit time of carriers across the device is much longer than the time scale of interest (the "long device" limit). For the case of interest here electrons moving at  $v_e = 1 \times 10^7 \text{ cm/s}$  cross a  $10\text{-}\mu\text{m}$  gap in 100 ps. Only the first 3 ps of transport is considered here so the assumption is well justified.

The contribution to the photocurrent due to the holes is considerably less than that for the electrons. The hole effective mass is much larger than the electron mass, hence the hole velocities are much smaller. A steady-state calculation of hole velocities by Brennan and Hess<sup>32</sup> indicates that the equilibrium hole velocity at  $E = 10 \text{ kV/cm}$  is only  $v_h = 2 \times 10^6 \text{ cm/s}$ , a factor of 5-10 times smaller than the electron velocity. The transient hole velocity is also expected to experience overshoot<sup>33</sup> but the degree of overshoot would have to be 5:1 or greater to compete with the electron overshoot. There is no reason to expect the hole overshoot to be this large. In addition, the hole overshoot takes place on a faster time scale because the density of states is higher and

hence scattering rates are larger than for the electrons. Thus any hole transients are expected to have subsided within the first 100 fs. For these reasons hole velocities have been neglected in the following calculations.

A calculation of the electron drift velocity and the relative populations in the  $\Gamma$  and L valleys for  $\lambda_{ex} = 620$  nm and  $E = 2$  kV/cm is shown in Fig. 14. At this photoexcitation wavelength the electrons have an excess energy of 519 meV, well above the threshold for  $\Gamma$ -L scattering (290 meV). Intervalley scattering is very efficient so that in the first 100 fs, 20% of the electrons have already scattered into the L valley. The electrons in the  $\Gamma$  valley relax down toward the bandedge, but as long as their energy is above 290 meV they continue to scatter into the L valley. After 500 fs the energy of the  $\Gamma$ -valley electrons has fallen below the intervalley scattering threshold, and the population in the  $\Gamma$  valley begins to rise again as electrons that initially scattered to the L valley begin to scatter back into the  $\Gamma$  valley. At equilibrium about 90% of the electrons reside in the  $\Gamma$  valley.

This behavior is reflected in the net drift velocity of the electrons, also shown in Fig. 14. The drift velocity follows the velocity of the  $\Gamma$ -valley electrons with only a small contribution from the L-valley electrons (at no time are there more than 25% of the electrons in the L valley). Hence

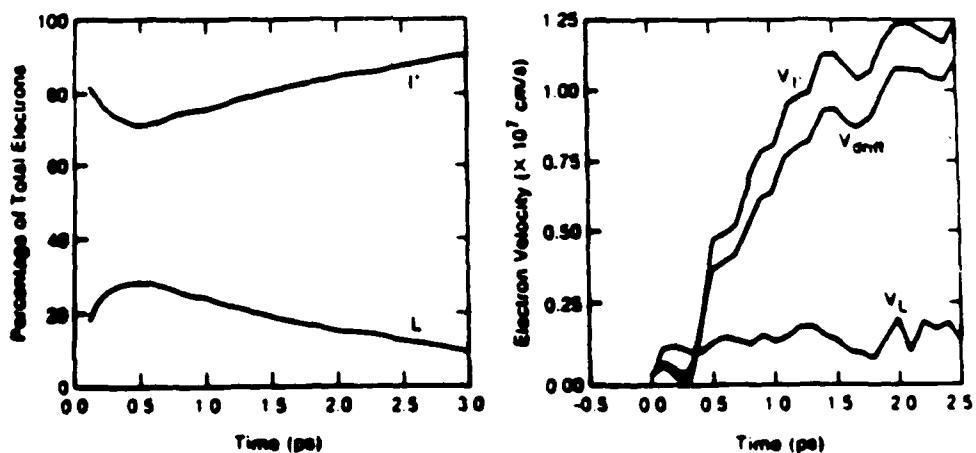


Figure 14: Transient  $\Gamma$  and L valley populations and velocity calculations for  $\lambda_{ex} = 620$  nm and  $E = 2$  kV/cm.

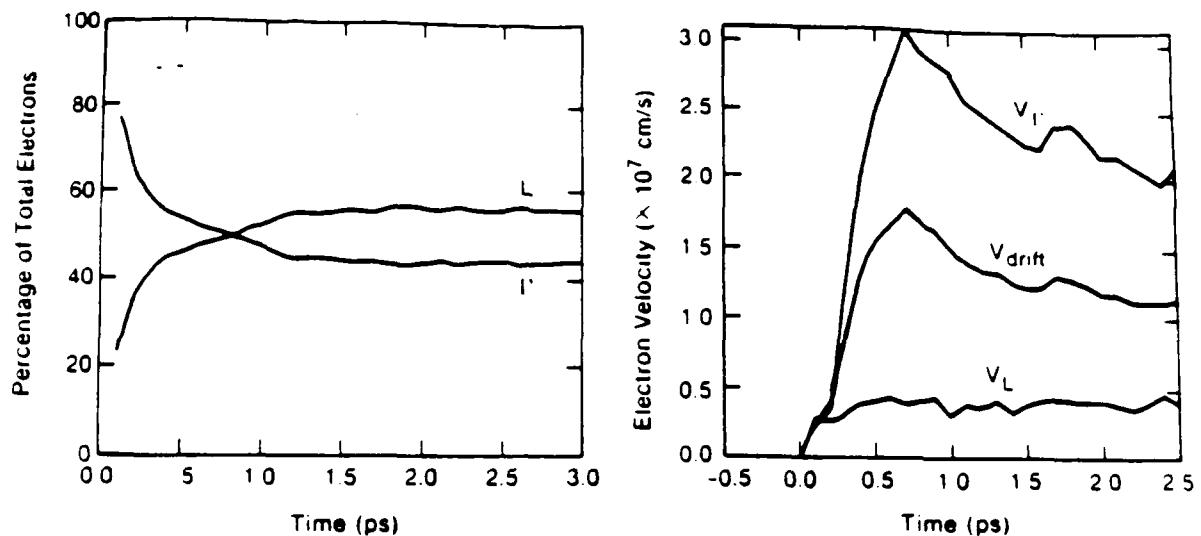


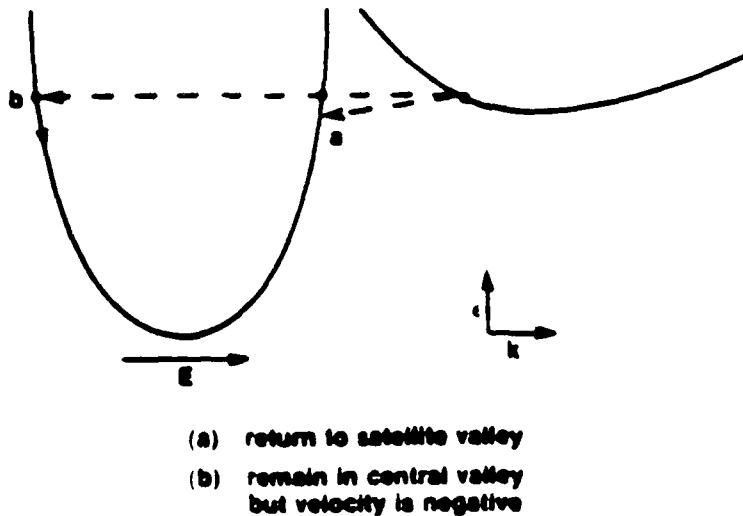
Figure 15: Transient  $\Gamma$  and L valley populations and velocity calculations for  $\lambda_{\text{ex}} = 620$  nm and  $E = 10$  kV/cm.

for this photoexcitation wavelength and applied field the electron drift velocity is expected to rise gradually to a constant value in approximately 2 ps with no velocity overshoot.

The situation is quite different for the high-field case shown in Fig. 15. In this case the relaxation of the photoexcited electrons toward the bandedge competes with the electron heating from the applied field. More electrons remain above the intervalley transfer threshold and hence more get scattered into the L valley. After 1 ps more than 50% of the electrons have scattered into the L valley, and at equilibrium 55% of the electrons are in the L valley and 45% reside in the  $\Gamma$  valley. Therefore the equilibrium drift velocity is the average of the velocities in the two valleys and is significantly reduced. This large percentage of electron transfer to the sideband and the subsequent slowing of electrons in the L valley due to lower mobilities is the source of velocity overshoot in GaAs. The degree of overshoot in this case is only 1.5:1 because electrons were initially above the intervalley transfer threshold and began scattering out of the  $\Gamma$  valley immediately. It will be shown that the closer electrons are photoexcited to the bandedge, the more dramatic will be the degree of velocity overshoot.

Consider again the low-field calculations of Fig. 14. One anomalous detail of the transient drift velocity is the delay in onset during the first 300 fs. At first this feature was thought to be insignificant and due to numerical noise, but it was found repeatedly for several different low-field calculations. In fact it is a signature of the Jones-Rees effect<sup>34</sup> and its potential role in transient photoconductivity measurements has been discussed by Grondin and Kann<sup>35</sup>. The phenomenon is shown schematically in Fig. 16. At  $t=0$  states (a) and (b) are equally populated by the laser excitation. Electrons initially at point (b) will cool down toward the bandedge and will have a negative velocity. Electrons generated at point (a) will be accelerated by the field and will scatter into the L valley and slow down. Therefore at low fields and short times the electrons will either have a negative or a small positive velocity and the net drift velocity will be close to zero.

For high fields, as already discussed above, many more electrons scatter into the L valley and the initial even distribution between states (a) and (b) is randomized very quickly. This is shown more clearly in Fig. 17 in which the total number of electrons with negative velocity is plotted for the first picosecond of transport. The contribution from negative-velocity electrons is lower at high field and therefore the drift velocity rises without any delay. Hence the prediction of



D. Jones and M. D. Rees, *J. Phys. C* 6, 1781 (1973).

Figure 16: Schematic diagram of the Jones-Rees effect.<sup>34</sup>

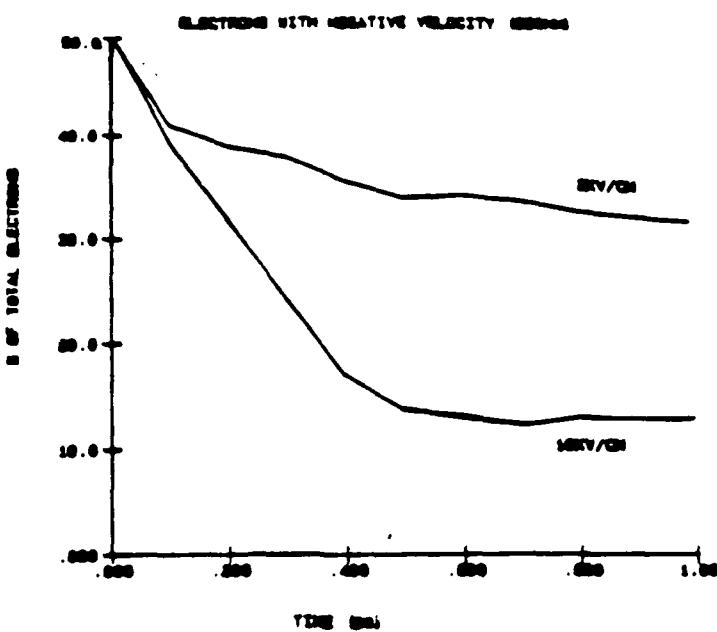


Figure 17: Calculation of the total number of electrons with negative velocity for conditions of low and high fields.

the Monte Carlo theory is that there will be a delay in the onset of the photoconductivity for low field relative to the onset at high field.

Most of the calculations of velocity overshoot published to date have assumed that the electrons were initially at the bandedge, and the degree of overshoot predicted was on the order of 3:1. This initial condition is the optimal condition for overshoot of the electrons before they gain enough energy to scatter to the sideband. For reasons to be discussed in the next section the experiment could not be performed with bandedge photoexcitation (871 nm) but could be performed at an intermediate wavelength of  $\lambda_{ex} = 760$  nm. In this case the electrons have an excess energy of 188 meV, well below the threshold for intervalley transfer. Monte Carlo drift velocity calculations for  $\lambda_{ex} = 760$  nm and several different applied fields are shown in Fig. 18. Note that, as expected, for  $E = 10$  kV/cm the degree of overshoot is enhanced to 2.5:1 and for the extreme field of 15 kV/cm the overshoot is 3:1.

This section has outlined the Monte Carlo simulations of the transient electron drift velocity that have been carried out for the experimental conditions of interest in this work. The following section will describe the transient photoconductivity experiments that have been performed and will compare these predictions with the experimental results.

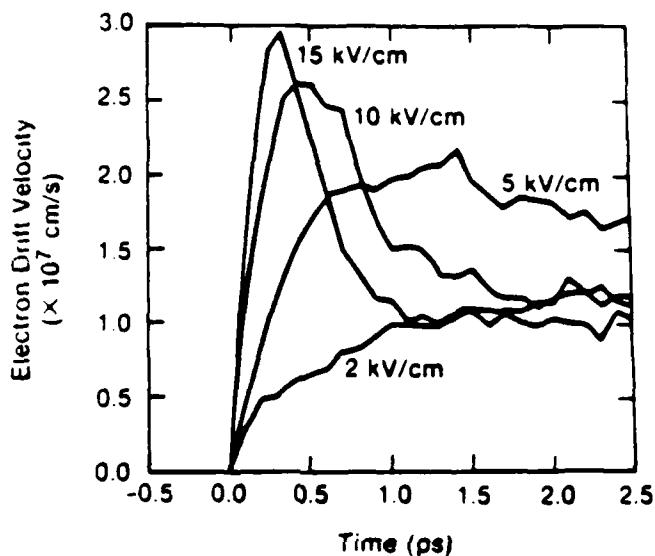


Figure 18: Predicted transient electron drift velocity for  $\lambda_{ex} = 760$  nm and several different applied fields.

### II.B.2. Experimental Observation of Velocity Overshoot and the Jones-Rees Effect

This section will detail the photoconductivity experiments that have been performed that constitute a time-resolved measurement of velocity overshoot and the Jones-Rees effect. Included in this discussion will be the development of different sampling geometries and the study of various forms of GaAs and contact metallizations to optimize the magnitude and reproducibility of the overshoot effect. The basic device under test is a simple photoconductive switch, consisting of a sample of bulk GaAs with two ohmic contacts, which is incorporated into a high-speed transmission line. According to the analysis by Auston<sup>84</sup>, when a dc bias is applied and the switch is "closed" using a short pulse of light (which generates photocurrent), a transient voltage is

launched in both directions on the transmission line away from the switch. The transient voltage is equal to the convolution of the transient photoconductivity of the switch with the gap capacitance. written as <sup>38</sup>

$$\alpha(t) = \alpha_0 \{ 1 - f_e(t) - f_h(t) \} \quad (5)$$

where  $\alpha_0$  is the unperturbed (time-independent) coefficient and  $f_e$  and  $f_h$  are the respective electron drift velocities. Therefore, in a situation in which  $n_e$ ,  $n_h$ , and  $v_h$  are approximately constant and  $Z_0 C_g$  is small relative to the time scale of interest, the transient voltage follows the transient electron velocity. With proper design these conditions may be satisfied in a simple photoconductive switch.

Once the voltage waveform has been generated it can be measured by the electro-optic sampling technique <sup>22</sup>.

### II.B.2.1. Excitation at 620 nm

The initial set of measurements were carried out using the coplanar stripline geometry (Fig. 12). This geometry was chosen for its improved temporal resolution, as discussed in section 4.2. The fabrication of the structure is as follows. A number of GaAs and LiTaO<sub>3</sub> pieces are cut and polished to form plates of approximately the same dimensions (typically 10x10x0.5 mm). One edge of each piece is ground and polished very carefully to achieve maximum flatness and smoothness. These respective flat edges of the two materials are then butted together, and the two pieces are mounted on a flat glass plate using UV-cured cement; the two pieces are pressed together during the cure to insure a very close fit. Next, the two materials are ground and polished simultaneously to an optical finish to insure a smooth contour from one material to the other. Finally, photoresist patterns are formed photolithographically across the interface, metal is evaporated onto the pattern, and the transmission lines and photoconductive gaps are defined through liftoff to the excess photoresist and metal.

One of the requirements of this geometry is that bulk (boule-grown) GaAs be used because there is considerable grinding and polishing of the material during fabrication (a thin epitaxial layer would not survive this treatment). Because it is also necessary to use semi-insulating material (so that the dark current is small relative to the photocurrent) the two options are to use either Cr-doped Bridgeman material or undoped LEC GaAs. The first experiments utilized Cr-doped GaAs because it was relatively inexpensive and readily available. In this case no dependence of the transient photoconductivity of the applied electric field was observed and in particular no photocurrent overshoot was observed. This was presumably due to two reasons. The first is that Bridgeman material is of relatively poor quality and the Cr (and other) impurities act as scattering centers and traps that severely limit the mobility of the electrons. The other is that all metallizations tested (including evaporated Au, AuGeNi, and In, both annealed and unannealed) resulted in nonreproducible non-ohmic contacts. This was reflected in highly non-linear dc I-V characteristics.

The next step was to utilize undoped LEC GaAs. In this case the contacts were more reproducible and the best contacts were obtained using annealed evaporated In. Even in the best case, however, the dc I-V characteristics varied significantly from one photoconductive gap to the next. Best-case transient voltage waveforms for these samples are shown in Fig. 19. This was the first instance of an observed photocurrent overshoot that was present at high fields but not at low fields. Two difficulties were apparent during the experiment. First, the shape of the voltage transient was strongly dependent upon the illumination of the gap, which is an indication of poor contacts<sup>36</sup>. Secondly, the rise times of the transients were unexpectedly long, on the order of 1.5 ps, which can only partially be explained by the relatively large linewidths and line separation (50  $\mu$ m) of the geometry used. Some of this slow response is also likely due to contact effects.

Improvements in the temporal resolution would require using smaller stripline dimensions. However, it was found that the difficulty in fabricating continuous transmission lines across the GaAs/LiTaO<sub>3</sub> interface increased dramatically as the linewidth was reduced. Yields for 20- $\mu$ m linewidths were extremely low and for 10- $\mu$ m linewidths or smaller were essentially zero.

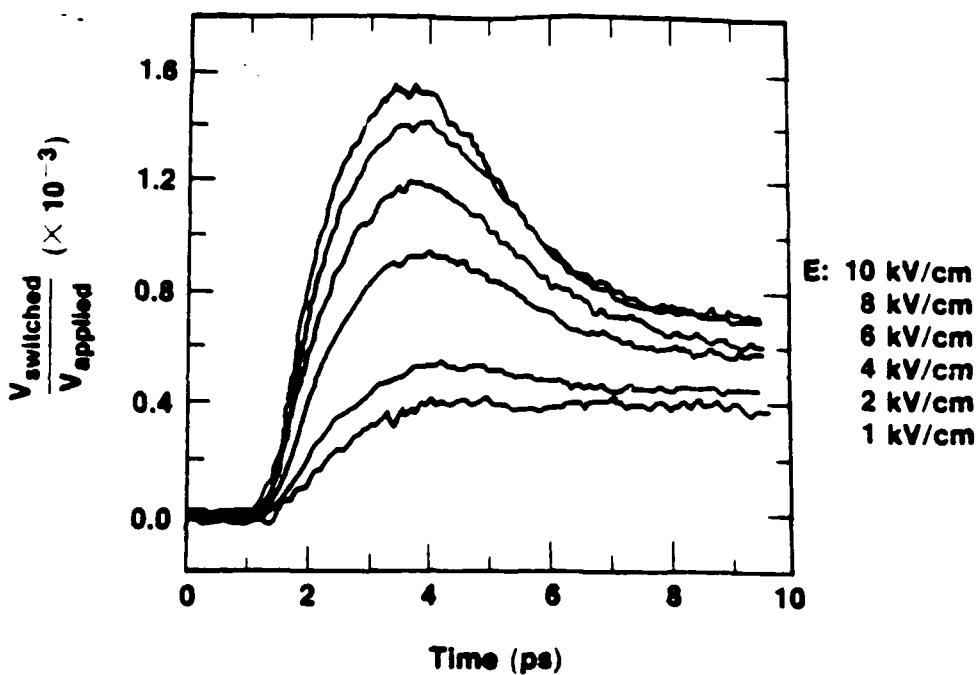


Figure 19: Transient voltage waveforms obtained using the coplanar strip geometry with semi-insulating LEC GaAs and annealed In contacts.

At this point it was decided to make two significant changes in the experiment. The first was to perform the experiment in the reflection-mode geometry and is shown in more detail in Fig. 20. This geometry has several advantages and one disadvantage. One advantage is that it separates the photoconductive switch fabrication from the sampling crystal fabrication, each of which can be done independently with standard procedures and very high yields. In particular, the coplanar strip linewidths on the GaAs may be made as small as practicable with the lithography at hand. Another advantage is that the probe beam may be moved arbitrarily close to the switch, thereby minimizing any distortion of the voltage waveform due to modal dispersion on the transmission line. The disadvantage is that the probe beam must pass through the fringing field twice which yields an effective optical transit time to that is twice that of the transmission geometry. Fortunately this is easily compensated for by sufficient reduction of the stripline dimensions.

The second change in the experiment is made possible by the transition to the reflection-mode geometry. Because grinding and polishing of the GaAs is no longer necessary, very high-quality MBE or MOCVD GaAs may be used which will have extremely low impurity concentrations and correspondingly high electron mobilities. The use of epitaxial GaAs has an additional advantage in terms of ohmic contact formation. Ohmic contacts require a heavily doped n<sup>+</sup> region between the metal and the semiconductor to serve as a source of electrons to the semiconductor. It is clear that the heavier and more uniform the doping the better the contact will be. Such a layer is easily incorporated into an epitaxial growth process as a cap layer and a very high-quality n<sup>+</sup>-n interface is assured.

As expected, it was found that the uniformity and quality of the contacts increased dramatically with the utilization of epitaxial GaAs. The best metallization was found to be sputtered NiAuGe annealed in an argon atmosphere at 400°C for 10-15 minutes.

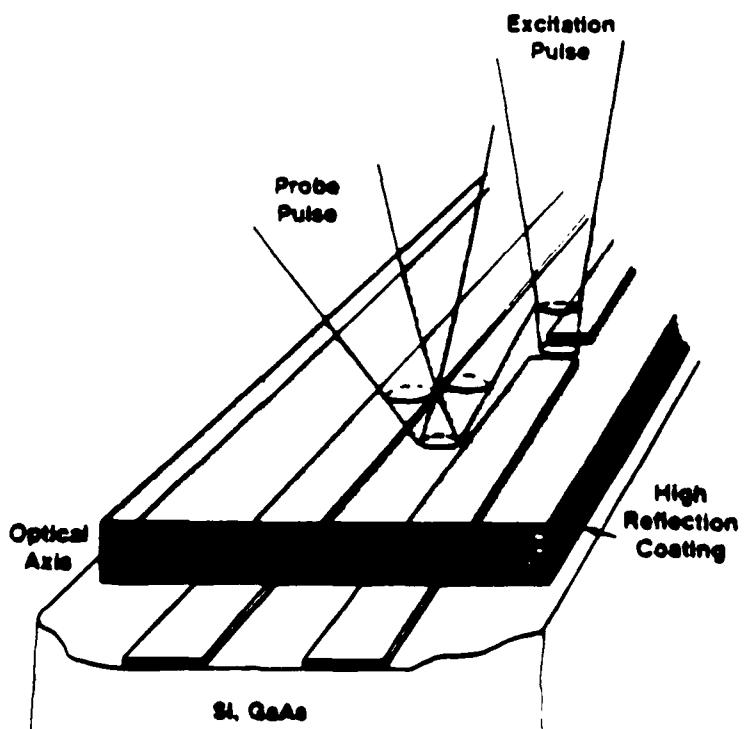


Figure 20: Reflection-mode electro-optic sampling geometry.

Experimental photoconductive transients for these samples are shown in Fig. 21.<sup>37</sup> For this measurement the coplanar linewidths and separation were each 10  $\mu\text{m}$ , the gap length was 10  $\mu\text{m}$ , the pump and probe beam diameters were 10  $\mu\text{m}$ , and the propagation distance from the gap to the probe point was 20  $\mu\text{m}$ . Although data was taken with uniform illumination of the gap, it was found that the shape of the voltage transient did not depend critically on how the gap was illuminated. The transients were reproducible with several samples from the same wafer as well as with samples from different wafers of both MBE and MOCVD GaAs.

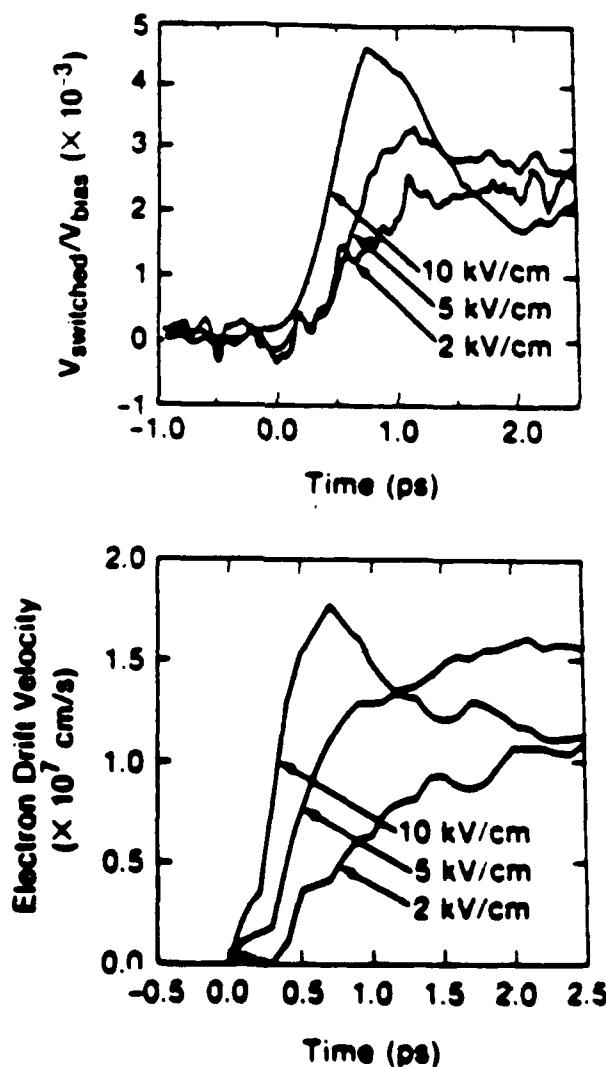


Figure 21: Transient photoconductivity data and Monte Carlo calculations for  $\lambda_{\text{ex}} = 620 \text{ nm}$ .<sup>37</sup>

Also shown in the figure are Monte Carlo calculations of the electron drift velocity for the particular experimental conditions of electric field, temperature (300 K), excitation wavelength (620 nm), and density ( $10^{17} \text{ cm}^{-3}$ ). All of the qualitative features of the theory are reproduced in the experimental results. These features include a decreasing rise time with increasing field and moderate overshoot that occurs at the extreme high field. Note also that the magnitude of the high-field overshoot prediction of 1.6 is in qualitative agreement with the measured overshoot of 2.1. It was shown that the temporal resolution for the sampling system with this geometry is under 200 fs. The measured response is a convolution of the transient drift velocity with the system response. Because the system response is considerably faster than the predicted velocity transient the measurement should reflect the inherent velocity response with only slight instrumental broadening. Therefore, because considerable care has been taken to optimize the temporal resolution and to fabricate reproducible ohmic contacts on high-quality epitaxial GaAs, the conclusion is that this measurement constitutes the first subpicosecond time-resolved observation of velocity overshoot in GaAs.

Another phenomenon predicted by the Monte Carlo theory and discussed in the previous section is the Jones-Rees effect<sup>34</sup>. The signature of this effect is a delay in the onset of the drift velocity for low fields when electrons are photoexcited at or above the intervalley transfer threshold. In this experiment this should be apparent as a delay in the onset of the photoconductivity at low fields. The effect should be observed for excitation wavelengths shorter than 700 nm.

Referring again to the data in Fig. 21 there is a shift that is apparent between the 10 kV/cm and 5 kV/cm curves. There is no visible shift between the 5 kV/cm and 2 kV/cm curves but this maybe obscured by the relatively large noise level on the low field data. One way of quantifying the onset time of the photoconductivity is by measuring how long it takes for the transient voltage to reach 10% of its peak value at each field. These measurements are shown in Table 1 and are compared with the relevant Monte Carlo predictions. For the high- and medium-field cases the

agreement is very good, and for the low-field case the agreement is within the uncertainty of the data and the calculation (low-field Monte Carlo calculations suffer from numerical instabilities). This constitutes the first time-resolved measurement of the Jones-Rees effect in GaAs.

	2kV/cm	5kV/cm	10kV/cm	
Monte Carlo Prediction		0.4 ± .05 ps	0.30 ps	0.10 ps
Photoconductivity Results		0.33 ± .05 ps	0.31 ps	0.12 ps

Table 1. Values of the measured photoconductivity onset times compared with the predictions of the Monte Carlo theory for onset of the electron drift velocity.

### II.B.2.2. Excitation at 720 nm

It was shown in the previous section that velocity overshoot increases as the photoexcitation energy approaches the bandgap because the electrons will spend a longer time in the central valley and therefore will have gained more energy from the field before scattering to the sideband. In an effort to measure an enhanced overshoot the transient photoconductivity experiment was repeated using the same optimized samples as above but with photoexcitation and probing accomplished using a near-IR dye laser running at 760 nm. Details of this laser source are found in the previous section. Note that for absolute optimization of velocity overshoot a short-pulse high-repetition-rate laser running at 850 nm should be used. At the time of this writing a laser with these properties has yet to be developed.

Transient photoconductivity data with 760-nm excitation at the high and low fields of 15 kV/cm and 2 kV/cm are shown in Fig. 22. Note that once again a very significant photocurrent overshoot is apparent at high field but not at low field. However, for the high field case the degree of overshoot is much smaller and the rise time of the photocurrent is considerably slower than what is predicted by the Monte Carlo theory (Fig. 18). In this case the limited temporal resolution is understood. The near-IR laser produces pulses only as short as 300 fs. Recall that the laser

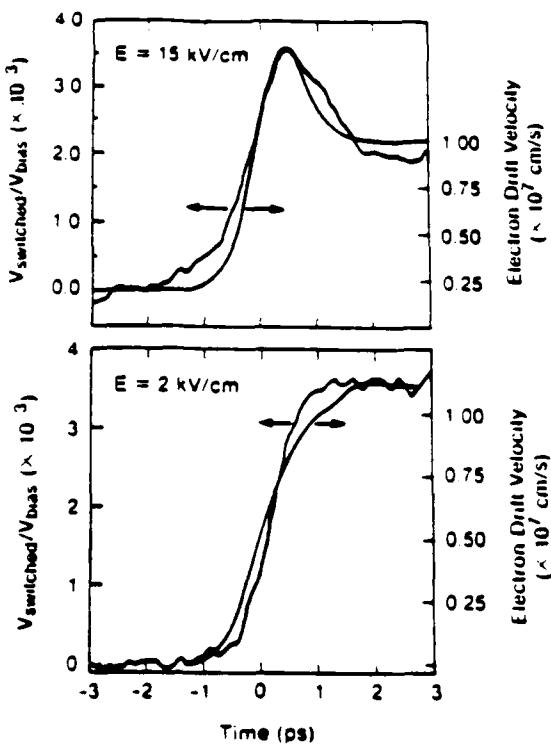


Figure 22: Transient photoconductivity results and Monte Carlo theory for  $\lambda_{ex} = 760$  nm. Note that the Monte Carlo curves have in this case been convolved with a 500 fs system response 37.

pulsewidth enters into the calculation of the temporal resolution twice, once for the probe pulse and once for the excitation pulse. In this case the non-negligible laser pulsewidth dominates the temporal resolution and therefore the full velocity overshoot cannot be resolved. Fortunately, since the temporal resolution is well known it is still possible to compare the data with the Monte Carlo predictions as long as the finite system response is taken into account. Specifically, if the theoretical predictions of Fig. 18 are convolved with the system response of 500 fs then a realistic estimate of the amount of overshoot that may be observed in this measurement is obtained. These curves are also plotted in Fig. 22. Once again very good agreement between the experimental results and the (convolved) theoretical predictions is found.

For this wavelength, as explained in the previous section, no Jones-Rees effect is expected. Unfortunately, the question of whether or not there is a delay in the onset of the photoconductivity cannot be answered because the temporal resolution is insufficient to resolve such a delay.

In this section the development of optimized samples, contacts, and sampling geometries to achieve a reproducible measurement of photocurrent overshoot has been described. Very good agreement has been obtained between transient photoconductivity results and Monte Carlo calculations of the transient electron velocity. In the next chapter an all-optical experiment will be described which is designed to measure not electron velocities but the electron distribution function. These two complementary approaches will yield a complete picture of the dynamic behavior of electrons in GaAs during heating by an applied electric field.

## II.C. Transient Absorption Spectroscopy

### II.C.1. Monte Carlo Model of the Dynamic Electron Distribution Function

The use of transient absorption techniques to measure the shape of the dynamic electron distribution function is based upon the bleaching of optical transmission due to the presence of excess carriers in the conduction and valence bands. The dynamic absorption coefficient may be written as<sup>38</sup>

$$\alpha(t) = \alpha_0 \{ 1 - f_e(t) - f_h(t) \} \quad (4)$$

where  $\alpha_0$  is the unperturbed (time-independent) coefficient and  $f_e$  and  $f_h$  are the respective electron and hole distribution functions. The transmission of a thin sample, neglecting interference effects, may be written as

$$T = (1 - R)^2 e^{-\alpha L} \quad (6)$$

where  $R$  is the reflection coefficient and  $L$  is the sample thickness. For thin samples ( $\alpha L \ll 1$ ) the exponential may be expanded and it is easily shown that the differential transmission has the form

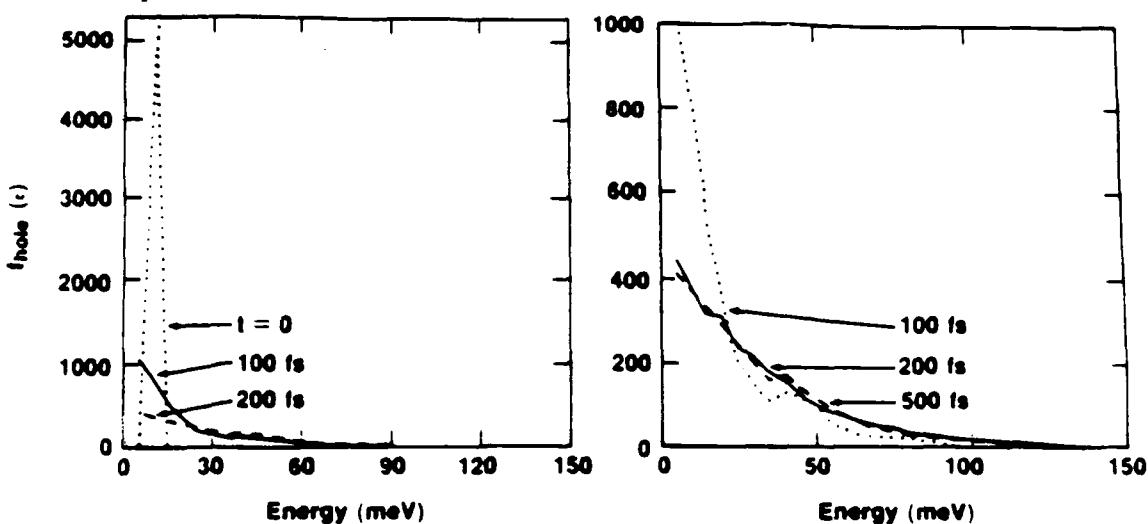
$$\frac{\Delta T}{T} = \frac{T - T_0}{T_0} \approx (\alpha_0 L)(f_e + f_h). \quad (7)$$

Therefore if the transmission data is plotted in this way the shape of the joint distribution function is obtained directly.

During velocity overshoot electrons are accelerated in the  $\Gamma$  valley under the applied field and eventually scatter into the  $L$  valley when they have gained sufficient energy. In terms of the energy distribution this implies that the distribution will develop a high-energy "tail". The measurement of the arrival time at high excess energies for electrons accelerated from bandedge amounts to a measurement of the scattering time into the  $L$  valley, which is of fundamental importance to velocity overshoot theories.

The Monte Carlo model is identical to that described earlier with a few modifications. Due to experimental limitations transient absorption experiments tend to be performed at higher densities than transient photoconductivity measurements (as high as  $n = 5 \times 10^{18} \text{ cm}^{-3}$ ), which requires the inclusion of electron-hole scattering effects. The finite spectral width of the excitation pulse (10 meV) has been included in the initial distribution. The excitation wavelength for the experiments is 830 nm, which generates electrons with an excess energy of 60 meV. This corresponds to approximately two LO phonon energies above the bandedge. The holes are initialized with an energy of 7 meV below the valence bandedge. The sample is assumed to be at  $T = 300 \text{ K}$  in all cases. The electron and hole distribution functions will be calculated as a function of pump/probe delay, carrier density, and applied field.

A calculation of the relaxation of the hole distribution with  $n = 10^{17} \text{ cm}^{-3}$  and  $E = 0$  is shown in Fig. 23. The very nonthermal initial distribution is randomized quickly and after 200 fs has evolved into a quasi-Fermi distribution. Note that all of the important dynamics take place



2472

Figure 23: Calculated evolution of the hole distribution with  $E = 0$  and  $n = 10^{17} \text{ cm}^{-3}$ .

within 50 meV of the bandedge. The experiment is focused on detecting changes in the tail of the distribution (above 100 meV), where the hole contribution is negligible, so the remainder of this discussion will be limited to the electron distribution.

Relaxation of the electron distribution function under the same conditions is shown in Fig. 24. At  $t = 100$  fs the distribution has two peaks, one at the initial photoexcitation energy and one at  $\epsilon = 30$  meV. The lower-energy peak is at exactly one LO phonon energy below the initial peak, indicating that LO phonon emission is the dominant scattering mechanism during the first 100 fs. After 200 fs both peaks have been smeared out by electron-electron and electron-hole scattering and the distribution is nearly thermalized. This 200-fs time scale for thermalization is consistent with calculation of Collet<sup>39</sup> and Bailey<sup>40</sup> discussed in the previous section.

A simulation for the same conditions but with a moderately high carrier density of  $n = 2 \times 10^{18} \text{ cm}^{-3}$  is shown in Fig. 25. The behavior is significantly different. In particular, the initial peak persists for up to 500 fs and no LO phonon peak is apparent. The persistent peak is

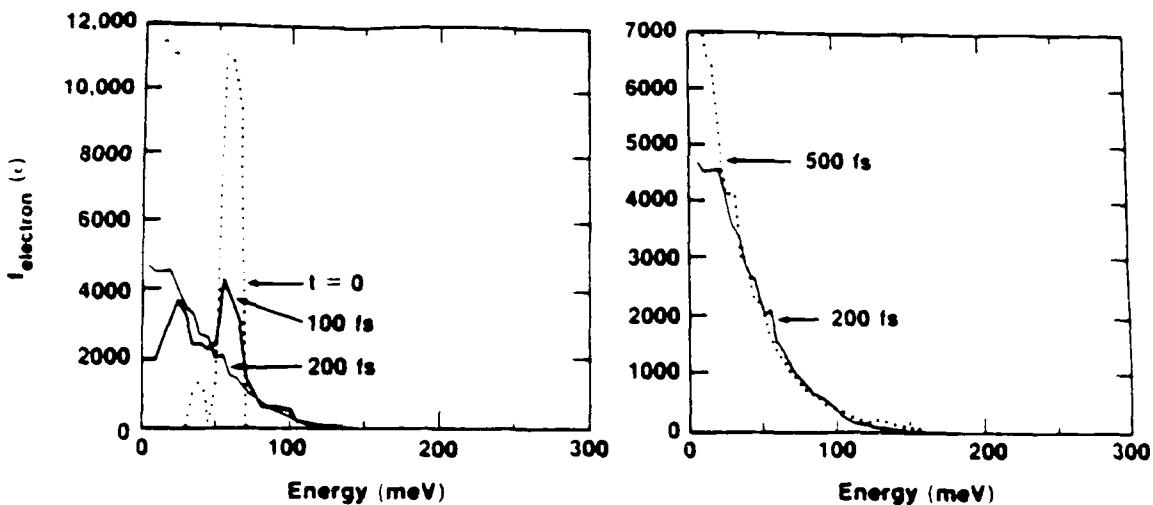


Figure 24: Simulated evolution of the electron distribution for  $E = 0$  and  $n = 10^{17} \text{ cm}^{-3}$ .

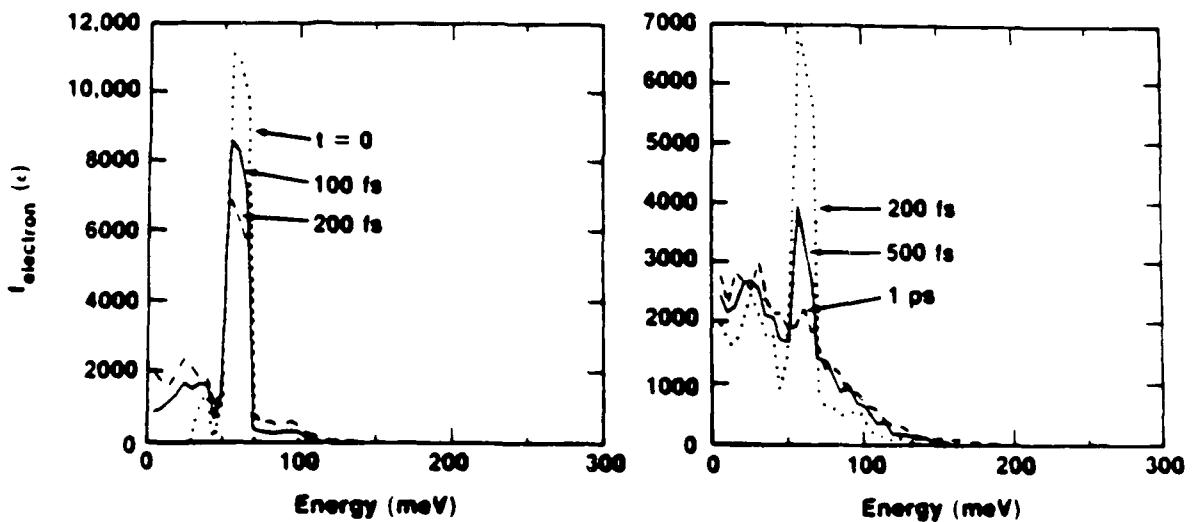


Figure 25: Calculated evolution of the electron distribution  $E = 0$ , and  $n = 2 \times 10^{18} \text{ cm}^{-3}$ .

due to two factors. First, at a density of  $2 \times 10^{18} \text{ cm}^{-3}$  the electron gas is moderately degenerate, and the scattering rate out of the initial state is limited by the number of states available to scatter into. Second, electron-hole scattering transfers energy to the hole distribution and prevents the immediate build-up of a high-energy tail. In this case the electrons are thermalized in approximately 1 ps. As will be discussed in the next section, no such long-lived nonthermal peak

has been observed in the experiments. The experiments were consistently conducted at as low a density as possible, estimated to be  $n = 2 \times 10^{18} \text{ cm}^{-3}$ . For these reasons the remainder of the discussion will be limited to densities on the order of  $10^{17} \text{ cm}^{-3}$ .

Next, a dc electric field is turned on in the simulation. Calculations for the high-field case of  $E = 10 \text{ kV/cm}$  are shown in Fig. 26. As expected, the applied field has a dramatic effect on the evolution of the distribution. A high-energy tail begins to form immediately and there is no LO-phonon peak. After 500 fs the tail continues to grow and significant numbers of electrons have energies as high as 300 meV, sufficient to scatter into the L valley (recall that for  $E = 0$  there were essentially no electrons with energies higher than 150 meV). Beyond  $t = 1 \text{ ps}$  the distribution continues to shift as electrons scatter back and forth between the  $\Gamma$  and L valleys, and equilibrium is reached after 2-3 ps.

A simulation for the same conditions but with a moderately high carrier density of  $n = 2 \times 10^{18} \text{ cm}^{-3}$  is shown in Fig. 25. The behavior is significantly different. In particular, the initial peak persists for up to 500 fs and no LO phonon peak is apparent. The persistent peak is

A different way to visualize the dynamics of the distribution function is to look at the distribution at discrete energies above the bandedge as a function of time. This corresponds to

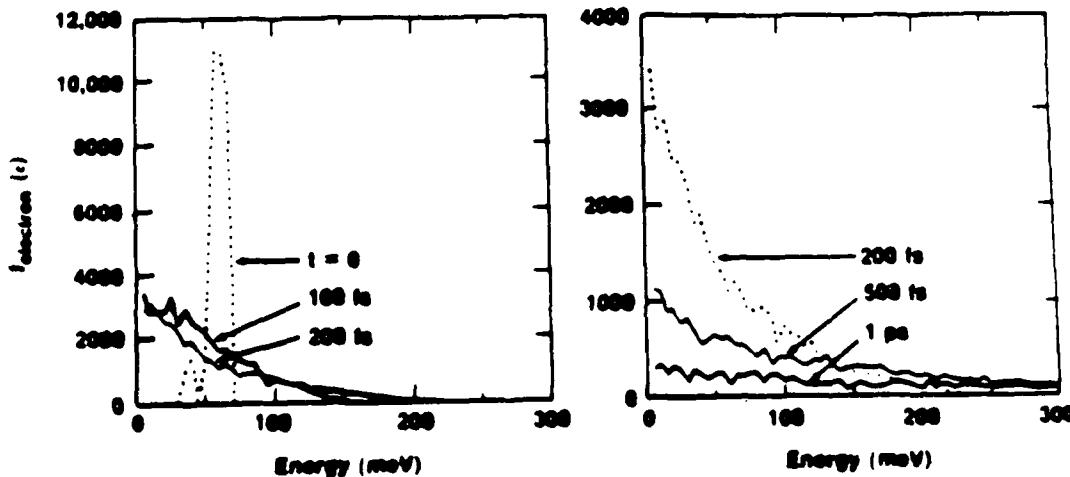


Figure 26: Simulated evolution of the electron distribution function, including heating by the applied electric field, for  $n = 10^{17} \text{ cm}^{-3}$  and  $E = 10 \text{ kV/cm}$ .

performing the experiment with discrete probe wavelengths rather than with a continuum probe, which will be discussed in the next section. Three probe wavelengths of interest are  $\lambda_{\text{probe}} = 870$  nm, 780 nm, and 750 nm, corresponding approximately to respective energies of 100, 150, and 200 meV above the bandedge. In the following plots the distribution function at these wavelengths will be weighted with the absorption coefficient ( $\alpha(100 \text{ meV}) = 10^4 \text{ cm}^{-1}$ ;  $\alpha(150 \text{ meV}) = 1.2 \times 10^4 \text{ cm}^{-1}$ ;  $\alpha(200 \text{ meV}) = 1.34 \times 10^4 \text{ cm}^{-1}$ ), in accordance with . (4) to allow for more direct comparison with the experimental results.

The data of Fig. 24 ( $n = 10^{17} \text{ cm}^{-3}$ ,  $E = 0$ ) is replotted in Fig. 27 for these three probe

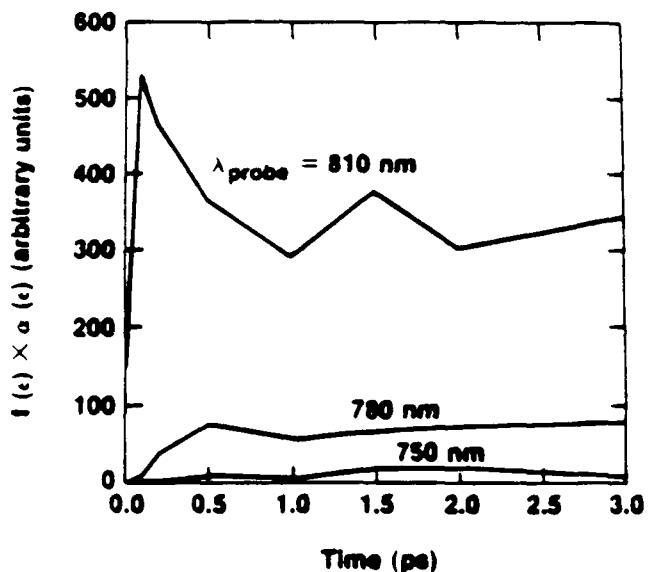


Figure 27: The calculations of Fig. 24 ( $n = 10^{17} \text{ cm}^{-3}$ ,  $E = 0$ ) are replotted for three discrete energies of interest as a function of pump/probe delay. The data is also weighted with the absorption coefficient.

wavelengths to demonstrate the advantage of this perspective. It can be seen immediately from this graph that there are fewer electrons at higher energies and that they arrive later. As already discussed, with no applied field there are extremely few electrons with energies as high as 200 meV.

Figure 28 simulates the same conditions but with the addition of the high applied electric field ( $E = 10 \text{ kV/cm}$ ). Two effects are apparent. A "population overshoot" occurs at all three

wavelengths as electrons are accelerated up from the bandedge, arrive at the probe energy, and sweep past to move to higher energies. Again a delay is seen as the arrival of the electrons is

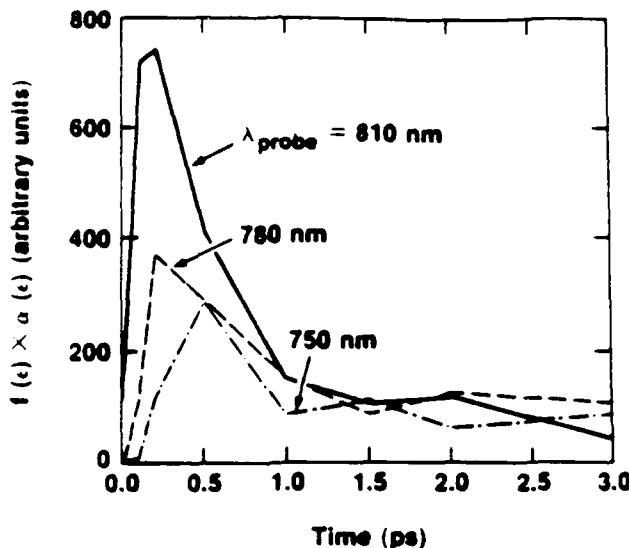


Figure 28: Evolution of the electron distribution function at discrete energies for  $n = 10^{17} \text{ cm}^{-3}$  and  $E = 10 \text{ kV/cm}$

probed further from the bandedge. Also note the very significant number of electrons in this case at  $\epsilon = 200 \text{ meV}$ . This is emphasized in Fig. 29, which compares the distribution function at 200 meV with and without the applied field. Essentially all of the electrons at this or higher energies have arrived there because of heating from the applied field.

In summary, Monte Carlo calculations predict that an applied electric field should have a dramatic effect on the evolution of the electron distribution function. A high-energy tail is expected to develop as electrons gain energy from the field. In particular, it should be possible to measure a transient absorption above 200 meV which is significantly enhanced when a high-electric field is applied.

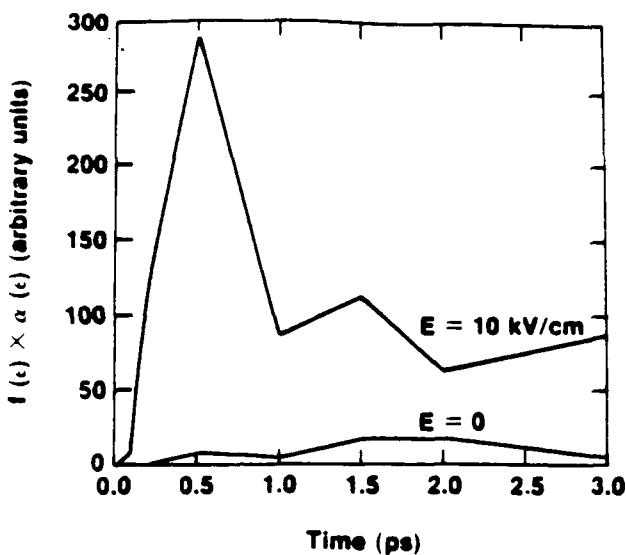


Figure 29: Comparison of the time dependence of the electron distribution function at  $e = 200$  meV for high field and zero field conditions.

### II.C.2. Experimental Results: Subpicosecond Heating and Thermalization of the Electron Distribution Function.

This section will describe the transient absorption experiments that have been performed with an applied electric field. The experiment will be described in two modes, with a continuum probe and with discrete probe wavelengths. The experimental results will be compared with the theoretical predictions of the previous section.

#### II.C.2.1. Continuum Probe Results

The dye laser and regenerative amplifier utilized to generate the white-light continuum for this experiment have been described in Section I. The output of the laser system is a train of high-peak-power pulses with energies of  $10 \mu\text{J}/\text{pulse}$ , a wavelength of  $620 \text{ nm}$ , a duration of  $100 \text{ fs}$ , and a repetition rate of  $1 \text{ kHz}$ . These pulses, when focused into a jet of ethylene glycol, water, or other medium, have sufficient peak power to generate a continuum through the high-intensity nonlinear index response of the material <sup>41</sup>. The discrete pump/continuum probe arrangement is shown in Fig. 30. The continuum maintains the original  $100\text{-fs}$  pulse width but the spectrum has

broadened to cover the range of 400-900 nm, which results in an extremely useful subpicosecond spectroscopic tool.

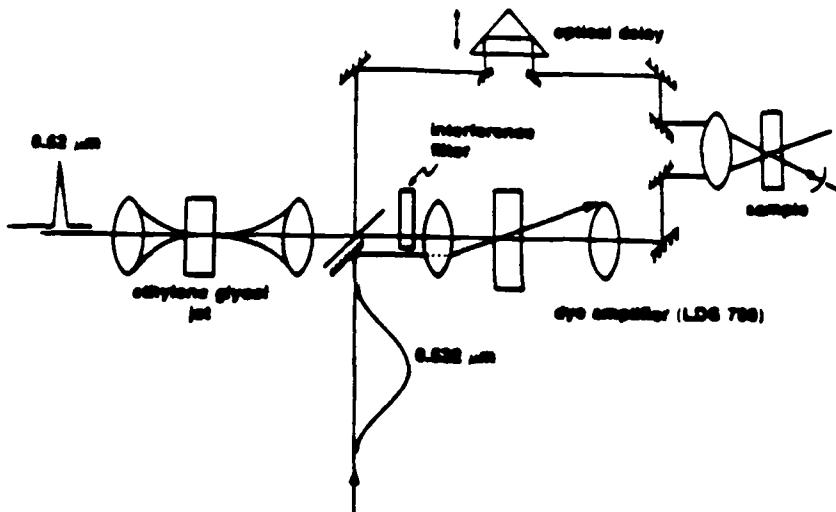


Figure 30: Discrete pump/continuum probe arrangement.

After the white light is generated it is split into two beams. The probe beam passes through a computer-controlled optical delay line, then is focused through the sample and the entire spectrum is recorded with an optical multichannel analyzer (OMA). Not shown in the figure is the compensation for the chirp on the probe pulse. Continuum generation with a relatively thick (1 mm) jet results in a pulse in which the red frequencies are leading the blue or vice versa, which is referred to as a "chirped pulse". In a femtosecond spectroscopy experiment all of the probe wavelengths must arrive at the same time. The chirp is corrected by passing the probe beam through a pair of high-index glass prisms<sup>42</sup>. In this way the chirp was reduced to 50 fs across the 700-850 nm band of interest, which is quite reasonable for a 100 fs pulse duration.

In the pump beam the pump wavelength of interest (in this case 830 nm) is selected out with a 10-nm bandwidth interference filter. The energy of the pump pulse is boosted with a single-

stage dye amplifier which is pumped with leftover light from the regenerative amplifier. The result is as much as 20 nJ per pulse in the pump beam. The pump energy was adjusted during the measurements to yield densities of approximately  $2 \times 10^{17} \text{ cm}^{-3}$ . The pump beam was aligned parallel with the probe beam and was focused into the sample through the same 50-mm lens.

The pump and probe pulses were timed by placing a second harmonic generation (SHG) crystal at the sample position and performing a cross-correlation measurement. In addition, it was necessary to carefully overlap the focused spots of the two beams at the sample position to optimize the transient absorption. This was accomplished by mounting a 12.5- $\mu\text{m}$  pinhole adjacent to the sample and scanning it both vertically and horizontally across the pump and probe spots. In this way the probe beam could be precisely centered over the pump beam. The probe beam diameter was typically 70  $\mu\text{m}$  and the pump diameter was 120  $\mu\text{m}$ .

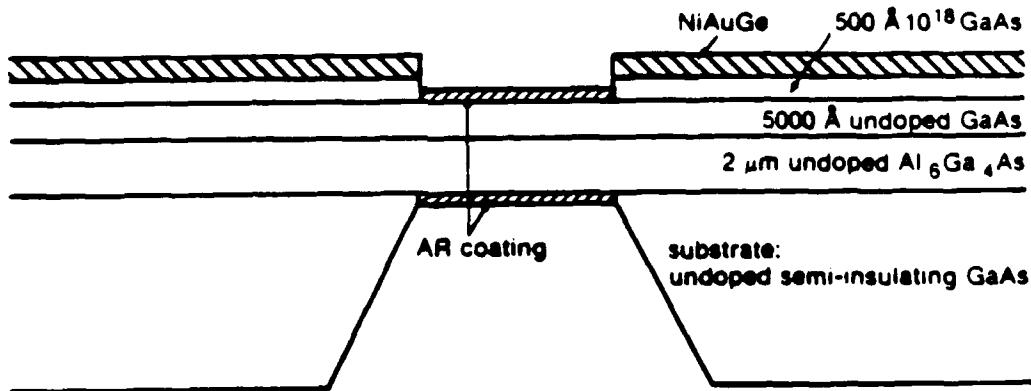


Figure 31: Sample structure for the transient absorption experiment.

The sample structure in these experiments is shown in Fig. 31. The MBE-grown layers consisted of 2  $\mu\text{m}$  of undoped AlGaAs (60% Al) grown on a semi-insulating substrate, followed by 5000 Å of undoped GaAs and a 500 Å  $n^+$  GaAs cap. The  $n^+$  cap was included to facilitate ohmic contact formation. The AlGaAs layer serves as a structural support for the thin GaAs layer and is optically transparent at the wavelengths of interest ( $\epsilon_{\text{gap}} = 2.1 \text{ eV}$ ). The metallization was annealed

NiAuGe. Windows were etched through the substrate using a  $H_2O_2/NH_4OH$  calibrated GaAs selective etch. The final fabrication step was to clean the samples thoroughly and then deposit onto both surfaces a thin dielectric broad-band anti-reflection coating. This reduced the reflection losses from 35% at each surface to less than 2% and helped to minimize interference effects. The metallic lines were 100- $\mu m$  wide and the length of the gap was 100  $\mu m$ . No attempt to fabricate a high-speed transmission line geometry was attempted as it was unnecessary for this experiment.

Experimental transient absorption curves for  $E = 0$  and  $n = 2 \times 10^{17} \text{ cm}^{-3}$  are shown in Fig. 32 for various pump/probe delays. The curves are plotted versus energy from the GaAs bandedge

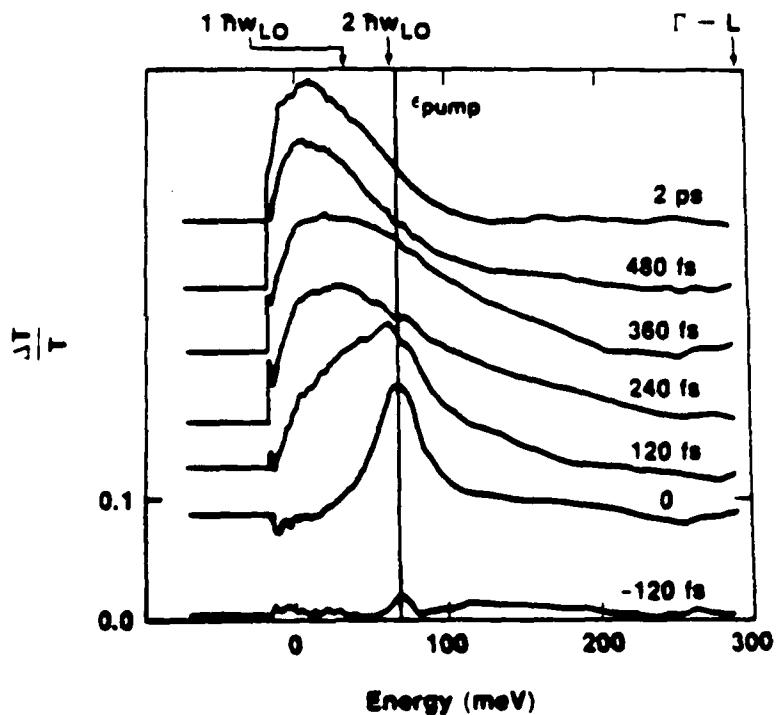
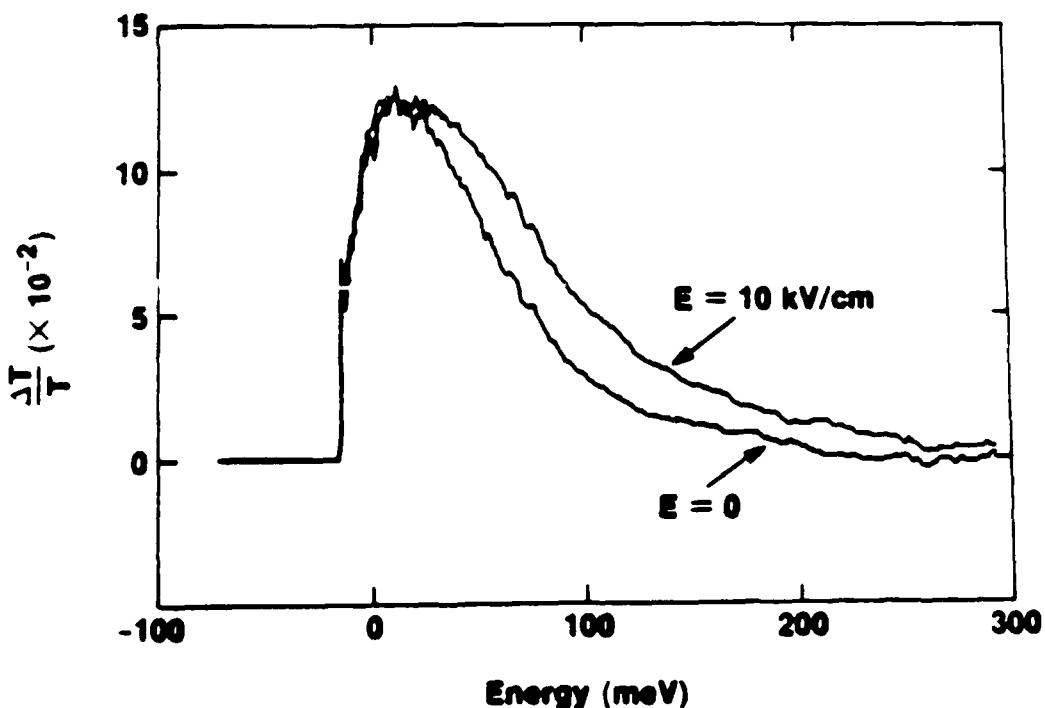


Figure 32: Experimental transient absorption data for  $E = 0$ ,  $n = 2 \times 10^{17} \text{ cm}^{-3}$ , and various pump/probe delays. The pump energy, LO phonon energies, and intervalley scattering threshold are indicated.

and each is offset vertically for clarity in the figure. Each curve required approximately 20 min of integration. The time origin is somewhat arbitrary and was chosen to coincide with the peak of the initial nonthermal distribution. Some contribution from the leading edge of the excitation pulse can be seen at  $t = -120$  fs. The distribution thermalizes rapidly and after 240 fs has become a quasi-

Fermi distribution. These results may be compared directly with the Monte Carlo predictions of Fig. 24. The presence of the LO-phonon peak at 120 fs is indicated by a shoulder in the experimental distribution at  $\epsilon = 30$  meV; this feature is not sharper because of the finite duration of the pump pulse (the Monte Carlo simulation does not take into account the temporal extent of the pump pulse). Note also that there is no evidence of a long-lived nonthermal peak such as is predicted for  $n = 2 \times 10^{18} \text{ cm}^{-3}$  (Fig. 25). The agreement between experiment and theory is very good for  $E = 0$  and a density on the order of  $10^{17} \text{ cm}^{-3}$ .

A measurement of the effect of a 10 kV/cm electric field for a fixed delay of 480 fs is shown in Fig. 33. As expected, the presence of the field enhances the transmission at shorter wavelengths, indicating a higher electron population in the tail of the distribution. Unfortunately it was found that the degree of this transmission enhancement was not very reproducible. Broad-band probe measurements are complicated by the fact that the shape of the white-light spectrum



2928

Figure 33: Comparison of the measured electron distribution at  $t = 480$  fs for  $E = 0$  and  $E = 10$  kV/cm.

tends to "wobble" even when all components have been optimized, which limits the sensitivity of the measurement to 5% changes in transmission. This is the same order as the expected changes in the tail of the distribution. For these reasons it was decided to proceed with the experiment using discrete probe wavelengths, a technique which improves the sensitivity by an order of magnitude.

### II.C.2.2. Discrete Probe Results

The principal of probing at discrete wavelengths is illustrated in Fig. 34 with the data of Fig. 32 used as an example. The three probe wavelengths selected are 810 nm, 780 nm, and 750 nm, corresponding to electron energies of 100, 150, and 200 meV respectively. As explained, with no applied field there are never very many electrons with energies as high as 200 meV, so only a very small transient absorption signal is expected at that energy. When a field is applied, however, electrons are heated by the field, a high-energy tail develops, and the transient absorption should be enhanced.

The simplest way to probe at discrete wavelengths would be to put interference filters in the probe beam in front of the sample. This is not an ideal situation because changing filters tends to misalign the pump/probe overlap and the optical delay introduced by each filter varies from one filter to another. This makes it difficult to make precise comparisons for different wavelengths. Instead, the full probe spectrum is focused through the sample and the discrete wavelengths are selected after the sample. One other modification is necessary, shown in Fig. 35. A beamsplitter placed in the probe beam divides it into two beams, one which passes through the sample and one which serves as a reference. The reference and probe beams are detected with matched spectrometers and PMT's. The spectrometers are set to the discrete wavelengths of interest. By taking the difference spectrum between the probe and reference beams and normalizing to the reference amplitude the pulse-to-pulse amplitude fluctuations in the white-light spectrum may be subtracted out, which yields an order of magnitude improvement in sensitivity. The sensitivity limit in this mode is a change in transmission of  $2 \times 10^{-3}$ .

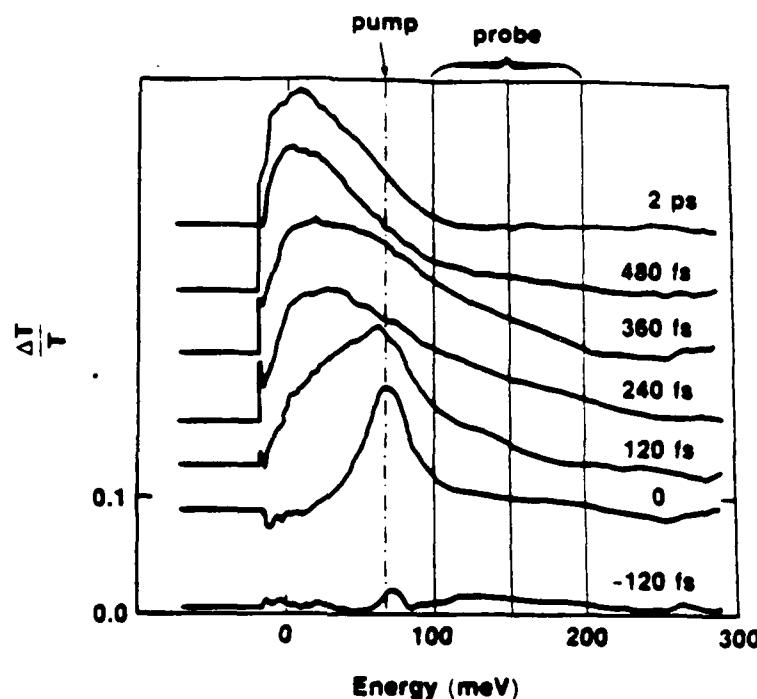


Figure 34: Discrete versus continuum probing of transient absorption.

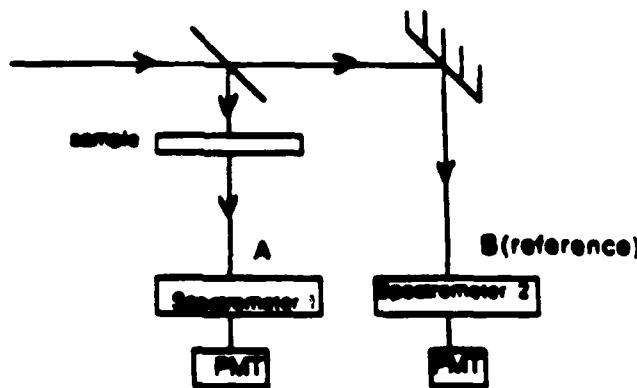
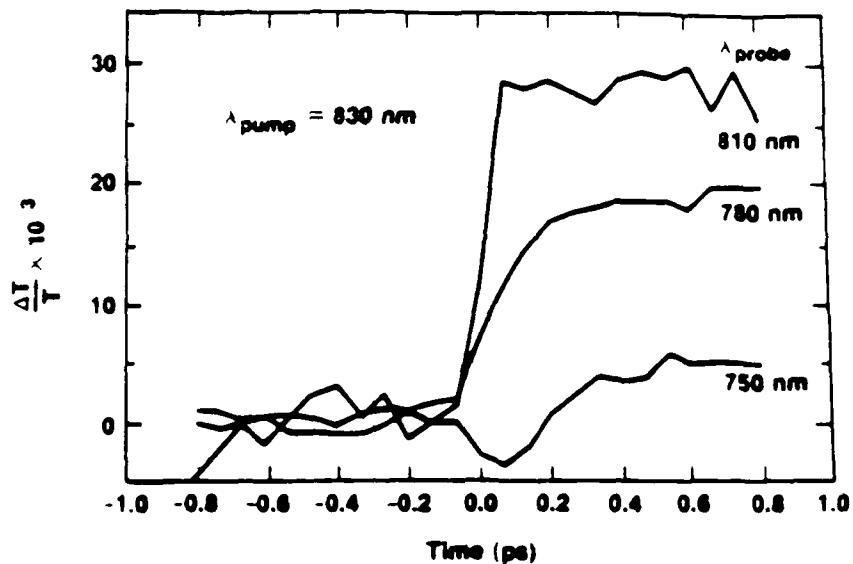


Figure 35: Details of the differential technique used in the discrete probe wavelength mode.

Experimental results for  $E = 0$  and  $n = 2 \times 10^{17} \text{ cm}^{-3}$  are shown in Fig. 36. The arrival of the electrons at  $\epsilon = 100 \text{ meV}$  is immediate, limited in this case by the pump pulselength. Electrons arrive later for  $\epsilon = 150 \text{ meV}$  and there are fewer at that energy. At the extreme energy of  $200 \text{ meV}$

they arrive later still and there are so few that the transient absorption is barely detectable. These results may be compared with the predictions of Fig. 34, all of the qualitative features of the theory have been reproduced in the measurement.



2493

Figure 36: Transient absorption data for three discrete wavelengths with  $E = 0$  and  $n = 2 \times 10^{17} \text{ cm}^{-3}$ .

The next step is to apply a dc bias to the sample. A 100 V bias across a 100- $\mu\text{m}$  gap should result in a uniform field in the sample of 10 kV/cm. The voltage switched across the gap was monitored on the load side of the gap with a high-speed oscilloscope. Under no conditions was the switched voltage greater than 0.5% of the applied voltage, indicating that the photoexcited density was low enough to induce only a small perturbation in the applied field. This justifies the assumption of a time-independent field in the Monte Carlo simulation. Transient absorption data for the intermediate probe wavelength of 780 nm ( $e = 150 \text{ meV}$ ) is shown in Fig. 37 for both  $E = 0$  and  $E = 10 \text{ kV/cm}$ . Clearly, there is an enhancement of the transmission with the applied field and the onset occurs earlier in time. This indicates that there are more electrons at this point in the band

and they arrive sooner with an applied field, as expected. However, there is no sign of the large "population overshoot" predicted for a 10 kV/cm field. This discrepancy will be discussed below.

Transient absorption data for the extreme wavelength of 750 nm ( $\epsilon = 200$  meV) is shown in Fig. 37 for  $E = 0$  and  $E = 10$  kV/cm. Although the signal-to-noise is not large in this

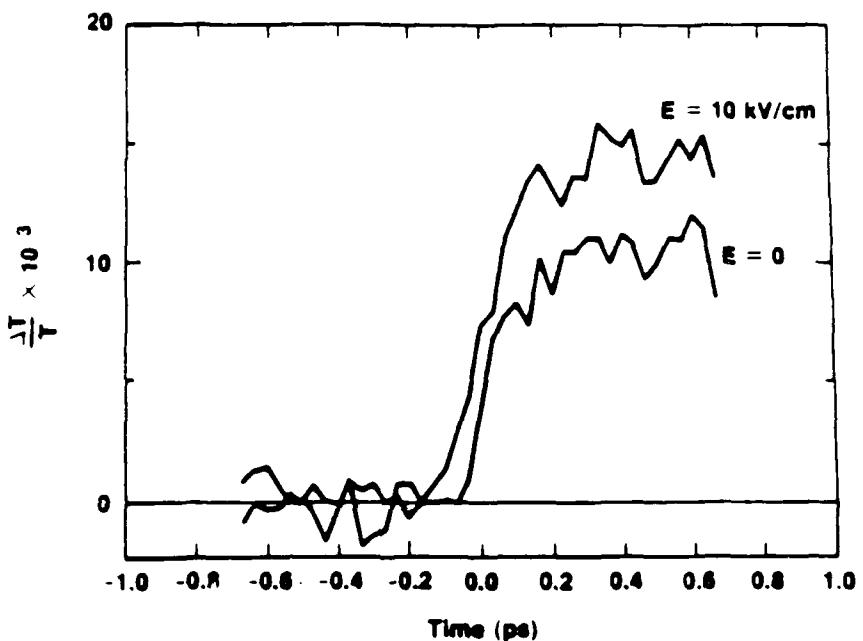
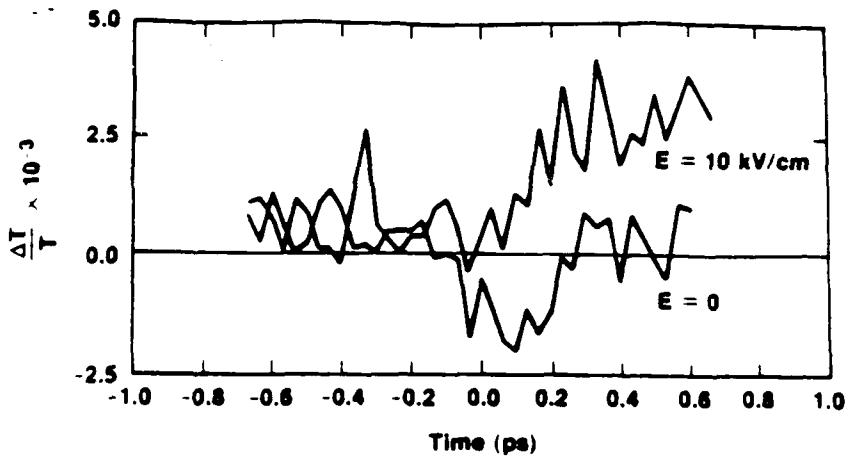


Figure 37: Transient absorption data with  $\lambda_{\text{probe}} = 780$  nm for high-field and zero-field conditions.

measurement, it is clear that there is an enhancement of the transmission when the field is applied. In fact there are so few electrons in this point in the tail of the distribution with  $E = 0$  that they cannot be detected. It may be said that a transient absorption signal was observed with the field on but could not be detected in the absence of the field, in qualitative agreement with the predictions of Fig. 36. Again the effect was not as large as predicted and no overshoot of the transmission was observed.

The data in Figs. 37 and 38, when compared with the Monte Carlo predictions, indicate the presence of a field but one which is much smaller than  $E = V_B/L$ . This could be due to poor contacts, but the contact fabrication used ( $n$ -GaAs/ $n^+$ -GaAs/AuGeNi) is essentially identical to that



2455

Figure 38: Transient absorption data with  $\lambda_{\text{probe}} = 750$  nm for high-field and zero-field conditions.

used for the transient photoconductivity samples. More likely is the formation of a stable high-field Gunn domain in the sample, which occurs when the background doping  $n_0$  or the sample length  $L$ , or both, are too large. When the domain forms, all of the applied field appears across the domain (at the anode) and the field in the rest of the gap is suppressed to a level below the threshold field of 3.8 kV/cm. The role of Gunn domain formation in GaAs photodetectors has been discussed by Wei *et.al.*<sup>43</sup> A linearized calculation by Jeppesen and Jeppsson<sup>44</sup> will be used to estimate the domain width for the samples parameters used here.

In Fig. 39 is plotted the domain width  $L_D$  relative to the gap length  $L$  for different applied biases and for representative values of the background doping. A polaron measurement of the doping versus depth for one of the absorption samples revealed that the doping level was  $2 \times 10^{18} \text{ cm}^{-3}$  near the surface (the  $n^+$  cap layer), fell to  $5 \times 10^{16} \text{ cm}^{-3}$  in the undoped n-GaAs layer, exhibited a spike up to  $1 \times 10^{18} \text{ cm}^{-3}$  at the GaAs/AlGaAs interface, and fell to  $5 \times 10^{16} \text{ cm}^{-3}$  in the AlGaAs layer. The spike may be due to contamination at the interface or formation of a two-dimensional electron gas. The relevant number is that in the GaAs and AlGaAs layers the

background doping is  $5 \times 10^{16} \text{ cm}^{-3}$ . Referring to Fig. 39 indicates that the extent of the domain under these conditions is 10-15% of the sample length. In other words, the high-field region is limited to the domain within 10-15  $\mu\text{m}$  of the anode; the field in the remainder of the sample is less than 3.8 kV/cm. This is consistent with the transient absorption data, which indicated the existence of a field but one considerably smaller than the applied field.

An attempt was made to focus the probe beam more tightly in order to sample the high-field domain close to the anode, but the smallest spot size attainable was 30  $\mu\text{m}$ , which was not small enough to resolve the domain.

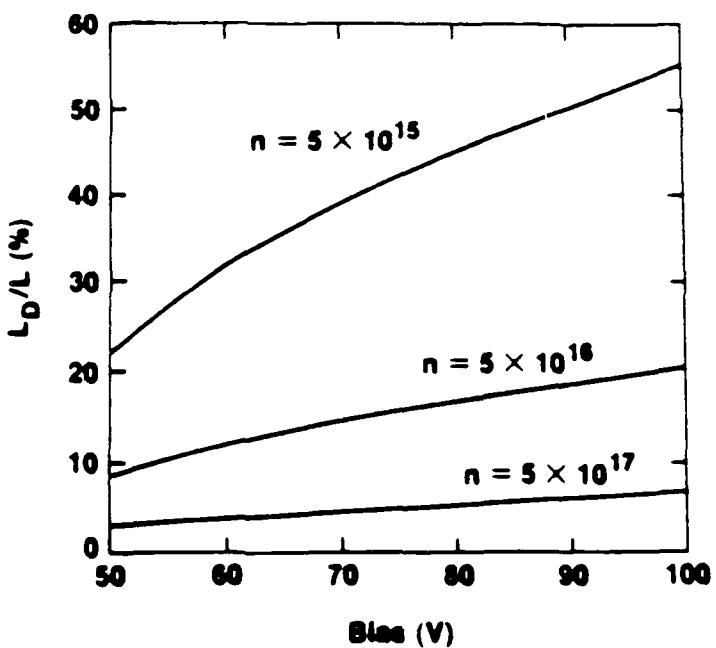


Figure 39: Calculation of the extent of the high-field domain versus bias for several representative doping levels.

There are several improvements in the experiment that could be carried out to defeat the domain formation. The primary solution is to reduce  $n_0 L$ .  $L$  is easily reduced by fabricating shorter gaps. The  $5 \times 10^{16} \text{ cm}^{-3}$  background doping in the GaAs layer is not ideal; that material can be grown with much lower background levels. In addition, the AlGaAs layer could be removed with a selective etch from the back side of the sample. This would eliminate the AlGaAs

contribution to the domain formation (the background level in the AlGaAs is about as low as can be expected with current MBE technology). Also it might be possible to pulse-bias the voltage on a time scale which is shorter than that required for domain formation but long enough to insure that displacement currents do not play a large role in the experiment.

The question arises as to why there was no indication of a nonuniform field in the transient photoconductivity measurements. In that case  $L$  was  $10 \mu\text{m}$  and the measured background level was  $5 \times 10^{14} \text{ cm}^{-3}$ . This yields an  $n_0L$  product of  $2 \times 10^{11} \text{ cm}^{-2}$ , which is below the threshold for stable domain formation estimated by Jeppeson and Jeppsson to be  $(n_0L)_{\text{threshold}} = 1.5 \times 10^{12} \text{ cm}^{-2}$ . In that case it is reasonable to assume that the internal field is uniform and given by  $E = V_B/L$ .

In this section experimental details of the transient absorption measurement have been described. Using the continuum probe approach relaxation of the initial nonthermal electron distribution on a 200 fs timescale has been observed, in good agreement with Monte Carlo predictions and results published in the literature. Some heating of the distribution due to an applied electric field was observed but was not extremely reproducible.

In the discrete-probe wavelength mode the experiment was repeated with an order of magnitude improvement in sensitivity. In this case the enhancement of the transient absorption was consistently observed with the application of the field, and in the high-energy tail of the distribution a transient signal was measured with a field applied that fell below the limits of sensitivity when the field was turned off. These results are in qualitative agreement with the Monte Carlo predictions, but indicate that the field in the sample is much lower than expected. A reasonable explanation of this discrepancy is that due to stable Gunn domain formation in the sample, most of the applied field appeared across the domain near the anode. Suggestions were described for future work to defeat the formation of the Gunn domain and perform the experiment with a uniform field.

### III. Tunneling in Quantum Well

#### III.A. Tunneling from a Single Quantum Well

We describe a set of experiments performed to investigate what might be called the "tunneling-out" problem in GaAs/AlGaAs quantum well (QW) structures. An electron is initially confined in a thin QW, where the well is coupled to a continuum through a thin barrier (Fig. 40).

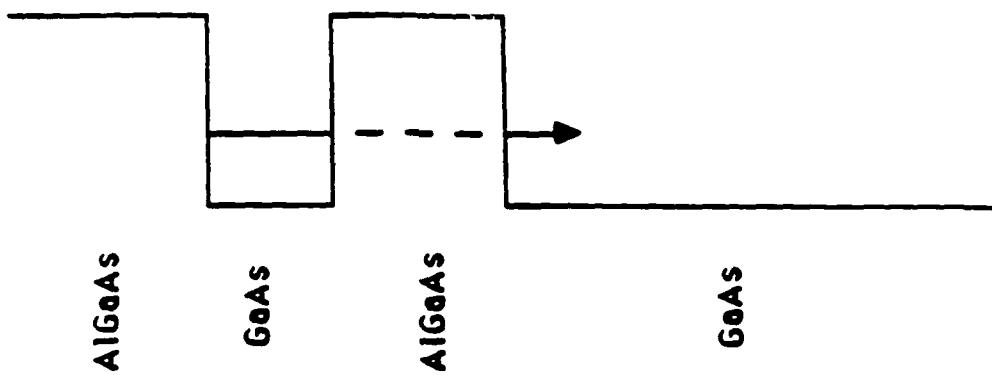


Figure 40: Conduction band diagram for the electron tunneling-out problem. The electron is initially confined in the QW, but decays via tunneling into the bulk GaAs region.

Hence the QW electronic state is really metastable, since the electron can tunnel out of the well into the continuum. This tunneling is irreversible, since after the electron escapes from the well it relaxes to the band edge of the continuum and hence cannot return to the well. The goal, of course, is to determine the tunneling decay rate of the electron, and the dependence of this rate on the barrier width, height, and an applied electric field. The tunneling of the heavy hole may be neglected; its effective mass is so large that the tunneling rate is orders of magnitude slower than that of the electrons.

The tunneling-out problem has been around in various forms for many years; among the famous manifestations of this problem are nuclear alpha decay and the cold emission of electrons from a metal surface (the so-called Fowler-Nordheim tunneling). The problem has been understood in principle since the early days of quantum mechanics; a comprehensive discussion of the theory of irreversible tunneling through a barrier has been given by Landau and Lifshitz.<sup>45</sup>

As was mentioned earlier semiconductor heterostructures offer a rich field for the study of tunneling in condensed-matter systems. In the GaAs/AlGaAs system, the barrier width may be controlled to monolayer tolerances, and the barrier height may be controlled by varying the aluminum composition of the AlGaAs barrier. Diode structures may be used to study the effects of an electric field. For this study, we have tuned the sample parameters so that the tunneling decay time falls within a range accessible by the time-resolved photoluminescence (PL) system described in the previous chapter. Thus we have been able to perform a fairly comprehensive study of the tunneling-out problem in QW structures using time-domain optical techniques. At the time we commenced this study, no such studies had been done. Recently, however, Tsuchiya, *et.al.*, published a similar study<sup>46</sup> in which they investigated the tunneling escape rate of electrons from a QW through very thin AlAs barriers. The principal differences between our study and theirs is that we used thicker AlGaAs barriers, where we could vary the Al composition to study the effect of varying the barrier height, and our samples had the QW region placed in a diode structure so that the effect of an electric field could be studied. We also had a single barrier instead of double barrier structure. For a double barrier structure, the electron can tunnel out of the well in either direction. Thus when an electric field is applied, the tunneling rate through one barrier increases, but the rate through the other barrier decreases. We used a single barrier structure so that the effect of the field on a single barrier could be isolated.

Aside from the question of principle regarding the nature of the tunneling-out problem, there is an additional, practical reason for studying this problem. The semiconductor heterostructure that has attracted the most interest for tunneling studies is the double-barrier diode (DBD). This structure is of particular interest for two reasons. First, from a practical point of

view, because they display negative differential resistance (NDR) with large peak-to-valley ratios. DBD's are useful for millimeter and submillimeter wave devices, such as oscillators and mixers. Oscillations at frequencies up to 56 GHz have been reported,<sup>47-49</sup> and detection and mixing at frequencies up to 2.5 THz.<sup>50</sup> Secondly, there are outstanding questions regarding the quantum-mechanical nature of the resonant tunneling in this structure. There are two related questions of particular interest. The first is, what is the time dependence (or conversely, the frequency response) of the tunneling in these structures? Second, what is the mechanism of the tunneling in real structures, where the effects of scattering must be considered? The latter question is relevant both to understanding the time dependence of the tunneling and the current-voltage (I-V) characteristics of DBD's.

Considerable theoretical effort has been directed towards answering these questions. The first calculations considered an ideal double-barrier structure with no scattering, interface disorder, or other mechanism by which the resonant energy level in the QW could be broadened. In this case (most thoroughly discussed by Ricco and Azbel<sup>51</sup>), the structure is completely analogous to a Fabry-Perot etalon, where one need only consider the transmission coefficients of each barrier to determine the I-V characteristic. The frequency response is determined by the requirement that for resonant enhancement of the tunneling, the electron probability amplitude must peak in the well. The buildup time for this peak is expected to be of the order of the resonant state lifetime.

In the fundamental paper, Luryi<sup>52</sup> pointed out that the Fabry-Perot picture of resonant tunneling fails to account for the observed frequency response and the I-V characteristics of real DBD's. The observed frequency response is much too fast and the I-V curves display smaller peak-to-valley ratios than predicted. He proposed an additional mechanism for resonant tunneling, namely "sequential" tunneling. The Fabry-Perot picture requires that the electron wavefunction be coherent over the entire DBD structure. However, any scattering process in the well will destroy this coherence. Hence, the tunneling process may be viewed as a two-step process, where the electron tunnels first into the well where it is localized by scattering, and then it tunnels out of the well. Many authors<sup>53-66</sup> have considered the time dependence of resonant tunneling when each

of these tunneling mechanisms is operating. Capasso, *et.al.*,<sup>67</sup> have reviewed the roles of coherent and incoherent tunneling in DBD structures. The important thing to note here, though, is that in either case, one of the most important parameters is the resonance width  $\Gamma_r$  of the confined electron state; whether the tunneling proceeds in a coherent or sequential fashion depends on the relative magnitude of  $\Gamma_r$  to the scattering width of the state  $\Gamma_s$ . The width  $\Gamma_r$  is related to the tunneling decay time  $\tau_T$  of the state by the Uncertainty Principle:  $\tau_T = \hbar/\Gamma_r$ . Our experiment directly measures in the time domain the dependence of this tunneling time and hence the resonance width on the applied electric field. Further discussion of these tunneling mechanisms may be found in Section IV.

### III.A.1. Experimental Results

#### III.A.1.1. Sample Structure

The structure of the samples used for this study is shown in Fig. 41. A single 30-Å GaAs QW is bounded on top by a thick (0.2  $\mu\text{m}$ )  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  barrier, and on the bottom by a thin  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  barrier and 0.1- $\mu\text{m}$ -thick GaAs region. This undoped structure is clad between and 0.1- $\mu\text{m}$  p<sup>+</sup> cap layer and a 1- $\mu\text{m}$  n<sup>+</sup> buffer layer grown on top of the semi-insulating (SI) substrate. It should be noted that the photon energy of the luminescence from the QW lines below

p <sup>+</sup>	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	2000
i	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	2000 Å
i	GaAs	30 Å
i	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	b
i	GaAs	1000 Å
n <sup>+</sup>	GaAs	1 $\mu\text{m}$
	S.I. Substrate	

Figure 41: Sample structure for the tunneling-out experiments

the  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  band edge, so there are no effects due to reabsorption of the luminescence from the QW.

Tunneling is possible only through the thin barrier beneath the QW. The barrier width  $b$  and Al composition  $x$  were varied to study the effect of these on the tunneling rate. In all, five samples were studied. One set of samples had an Al composition  $x$  of 30%, and nominal barrier widths  $b$  of 85, 111, and 121 Å. A second set had a barrier width  $b=86$  Å, and Al compositions  $x$  of nominally 38% and 50%. A 600-μm-diameter mesa was etched down to the  $n^+$  buffer and ohmic contacts were made on both  $n^+$  buffer and  $p^+$  cap, except a 200-μm-diameter hole was left for optical access<sup>68</sup>. The p-i-n diodes were operated in reverse bias. Typical current-voltage characteristics are shown in Fig. 42, with and without laser light incident on the sample. The reverse bias breakdown voltage was typically about -4V.

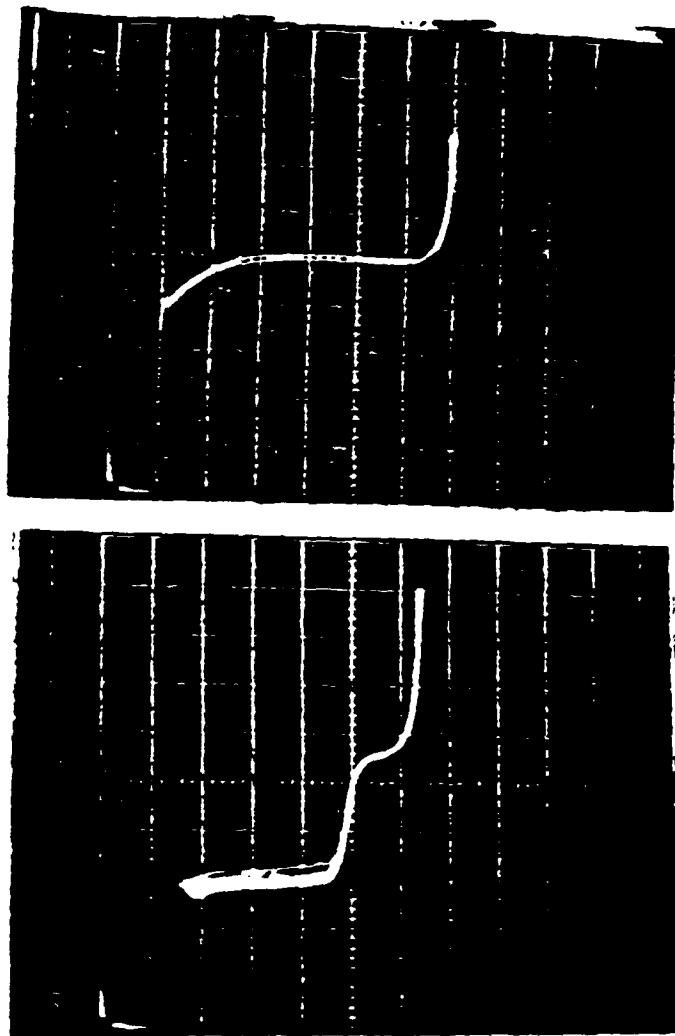


Figure 42: Current-voltage (I-V) characteristic for the sample shown in Fig. 2. On top is the I-V curve with no light incident on the sample. Below is the I-V curve with the laser pump beam incident. The horizontal scale is 1 V/division; the vertical scale is 20  $\mu$ A/div.

### III.A.1.2. Excitation and Luminescence

Electron-hole pairs were injected into the QW at  $t=0$  by a picosecond pulse from a dye laser. The laser system was described earlier. The laser dye was Pyridine 1; the laser was tuned so only electron-heavy hole pairs were generated. This was to avoid complications due to light hole tunneling through the barrier. However, we found that there was no difference in the results when light holes were also generated; evidently the light hole-heavy hole relaxation time is fast enough so the observed PL decay rates are unaffected.

The samples are held in a cryostat at a temperature of 6 K. At this temperature virtually all the electron-hole pairs are bound in excitons. The strong exciton luminescence spectrum was detected and time-resolved. A typical time-resolved PL spectrum is shown in Fig. 43. The advantage of performing time-resolved spectroscopy is that the Stark shift of the PL line with applied electric field can also be simultaneously measured, along with the time dependence of the luminescence. Furthermore, the scattered pump light is also accumulated on the spectrum, so the precise time origin is given.

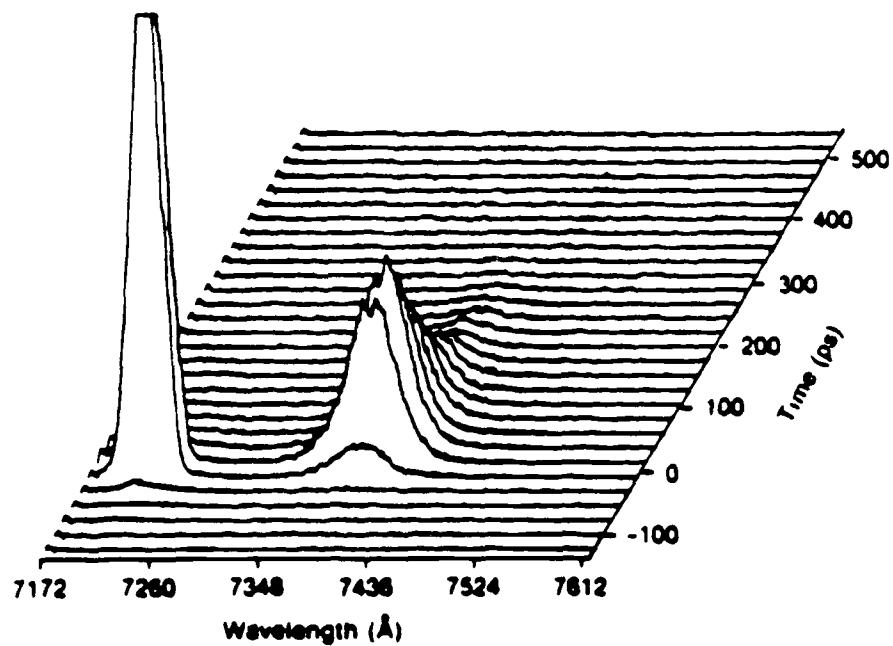


Figure 43 Typical time-resolved photoluminescence spectrum. On the left is the scattered laser pump light; on the right is the luminescence from the quantum well.

### III.A.1.3. PL Decay Times

The PL decay rate is expressed as

$$\frac{1}{t} = \frac{1}{t_T} + \frac{1}{t_R} \quad (7)$$

where  $\tau_R$  is the electron-hole recombination (radiative and nonradiative) time, which is of the order of a few hundred picoseconds. (The actual value depends on the density of trap levels and the injected carrier density). However, the escape time  $\tau_T$  is quite independent of those effects and would be equivalent to the measured PL decay time if it were much faster than  $\tau_R$ .

The PL decays were, therefore, fitted with single exponential decays. For decay times much longer than the streak camera response time of 20 ps, this was done by simply performing a least-squares fit on the PL decay at  $t > 20$  ps. For decay times close to the streak camera resolution, the data was fitted by using a simple rate equation model so that the effect of the rise time and finite streak camera resolution could be included. The model assumes that electrons are generated with a rate  $g$  (described by a Gaussian pulse with FWHM and time origin determined by the measured pump pulse) into the QW conduction band. The occupation of the initially populated state is described by  $n_1$ . The electrons relax to the conduction band origin, where they may bind with the holes to form excitons; this state is described by  $n_2$ . (The hole relaxation is assumed to be extremely fast<sup>69</sup>). Thus the combined subband relaxation and the exciton formation time are included in a rate  $k_{sr}$ . The electron tunneling rate is  $k_T$  and the recombination rate is  $k_R$ . Hence, the rate equations are

$$\frac{dn_1}{dt} = g - (k_T + k_{sr}) n_1 \quad (9)$$

$$\frac{dn_2}{dt} = k_{sr} n_1 - (k_T + k_R) n_2 \quad (10)$$

The observed luminescence is just proportional to  $n_2$ .

The equations were numerically integrated, and the rates  $k_{sr}$  and  $k_T$  varied to give a best fit. (The rate  $k_R$  is too small to be important for the case of fast tunneling where the rate equation

fit is required). An example of the fit is shown in Fig. 44, for the case of the sample with  $x = 0.3$ ,  $b = 85\text{\AA}$  at a bias voltage of  $-1.75\text{V}$ . A good fit is obtained for  $1/k_{\text{tr}} = 20 \text{ ps}$  and  $1/k_T = \tau_T = 31 (+ - 5) \text{ ps}$ .

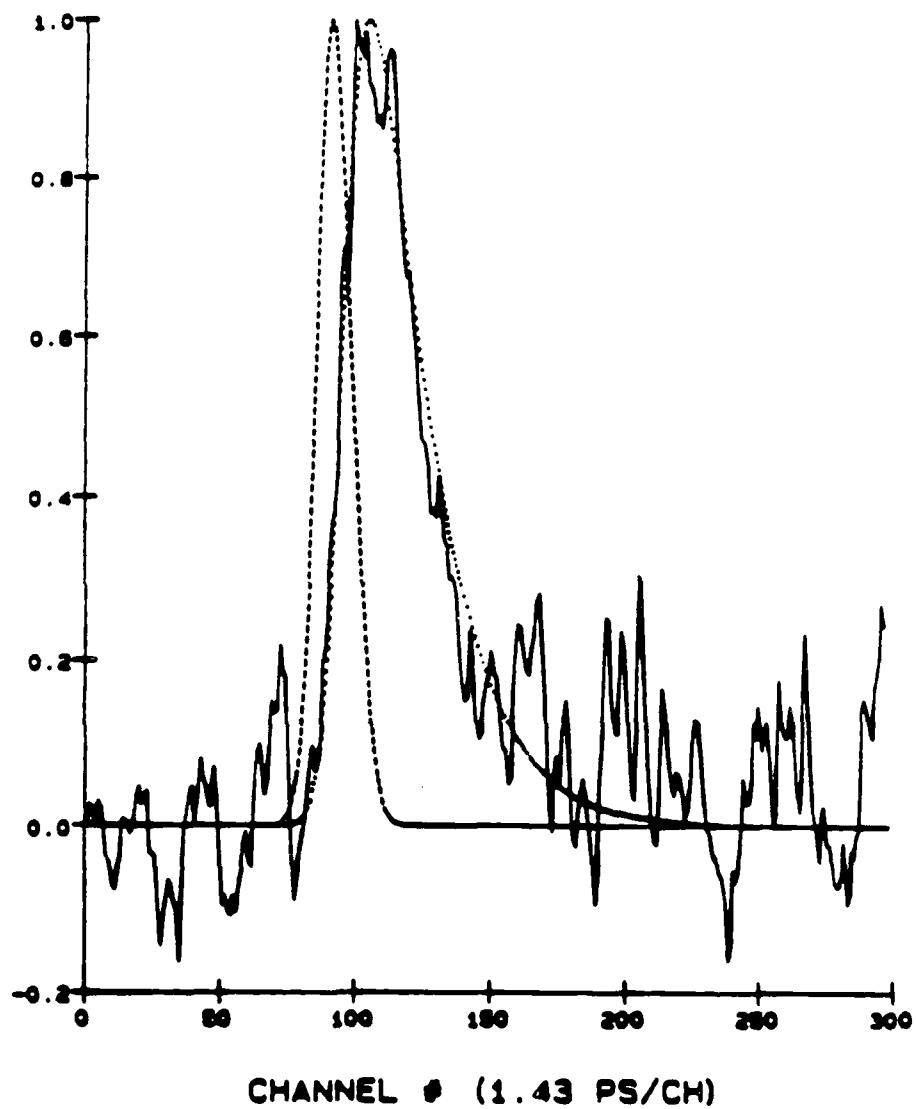


Figure 44: Rate equation fit of the luminescence data for the sample with Al composition  $x = 30\%$  and barrier width  $b = 85\text{\AA}$ , with an applied bias of  $-1.75\text{V}$ . The dashed line is the laser pulse, the dotted line is the solution for  $n_2$  from the rate equations, and the solid line is the experimental data. (The time axis is plotted in units of the detector channel number).

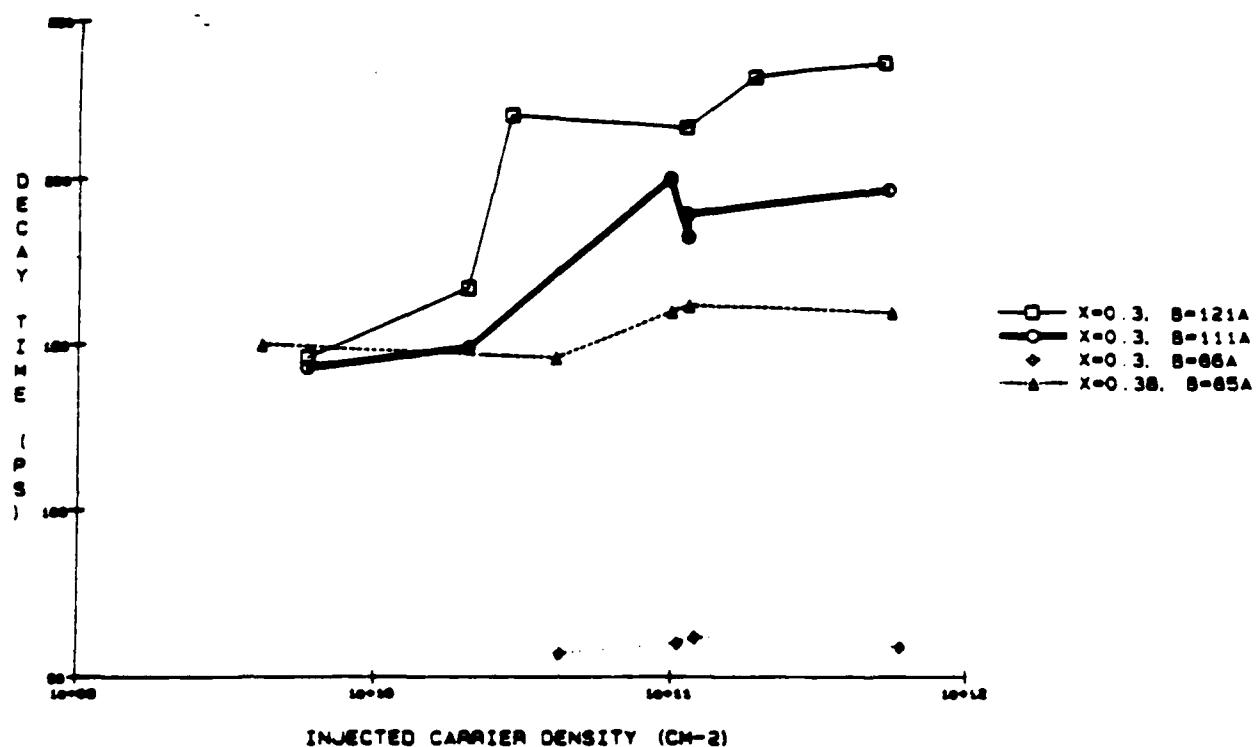


Figure 45: Luminescence decay time vs injected carrier density.

The decay time vs calculated injected carrier density are plotted in Fig. 45. There was no applied electric field for this data. The data shows a density-dependent decay time. At low density, the decay time for all samples is about 150 ps, except for the smallest barrier sample ( $x = 0.3, b = 85\text{\AA}$ ), which has a density-independent decay time of about 60 ps. For larger barrier samples, the decay time increases with density, though this decay time seems to saturate at high density. This behavior has been observed by other workers<sup>70</sup>, and is attributed to trap saturation. At low injected density, many carriers (probably holes<sup>70</sup>) fall into traps, thus increasing the observed PL decay rate. At high density, the traps are filled, so the observed PL decay is due only to recombination and tunneling.

### III.A.1.4. Field Dependence of PL Decay

The PL decay as a function of the applied bias is shown in Figs. 46 and 47 for the various samples. For the data in Fig. 46, the Al composition  $x$  is 30%, and the barrier width is varied. For Fig. 47 the barrier width is constant the Al composition is varied to study the effect of changing the barrier height. The estimated injected carrier density for these experiments is about  $2 \times 10^{10} \text{ cm}^{-2}$ . (The observed PL decay times are somewhat longer than observed in the experiment of Fig. 45. The discrepancy is likely due to the uncertainty in determining the laser spot size on the sample, and therefore the precise value of the carrier density.)

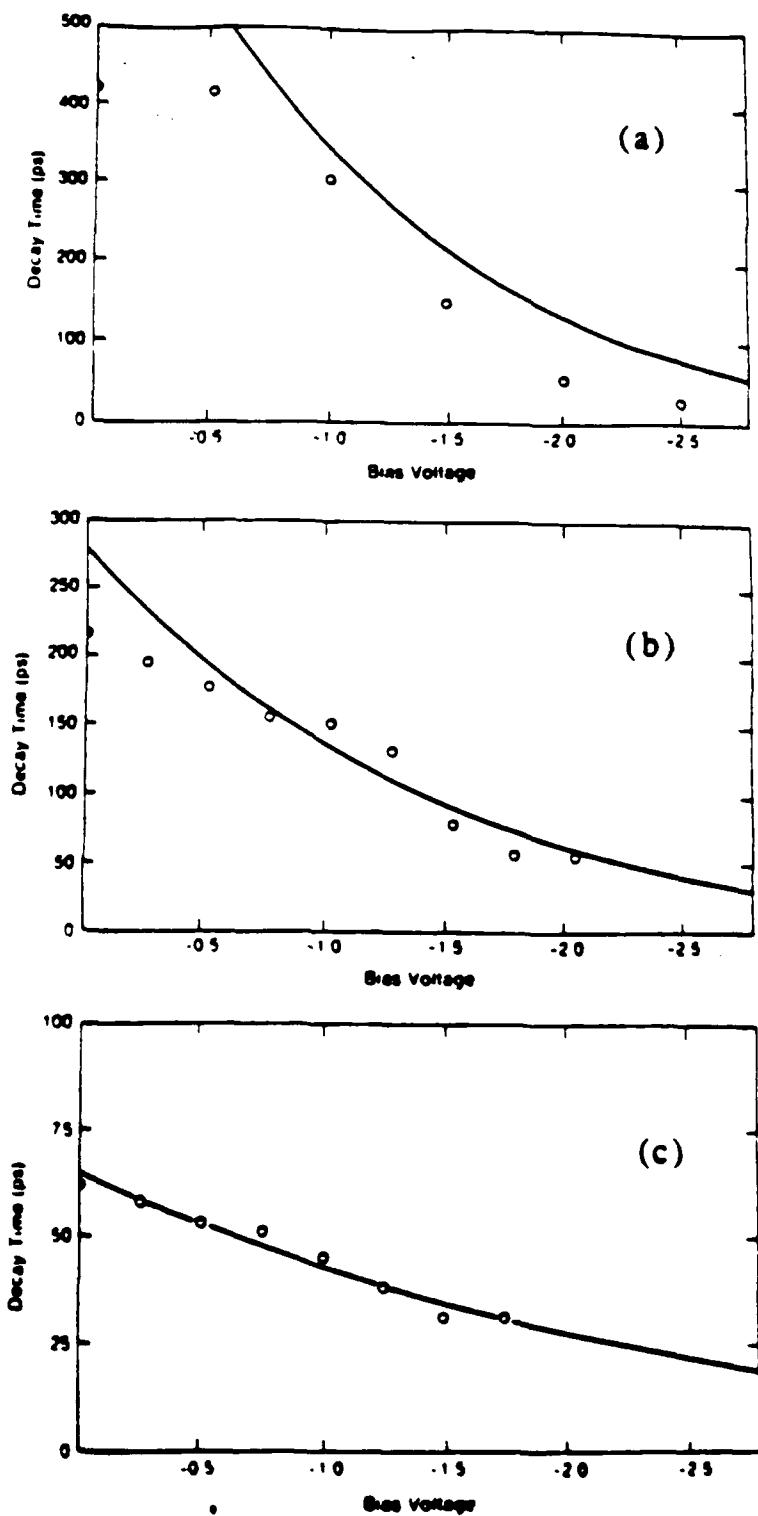


Figure 46: Luminescence decay time vs. applied bias for different barrier widths. The  $A_1$  composition  $x = 30\%$ . The barrier widths are (a)  $121\text{\AA}$ , (b)  $111\text{\AA}$ , and (c)  $86\text{\AA}$ . The solid lines are the results of the tunneling-out theory described in section III.C.3.

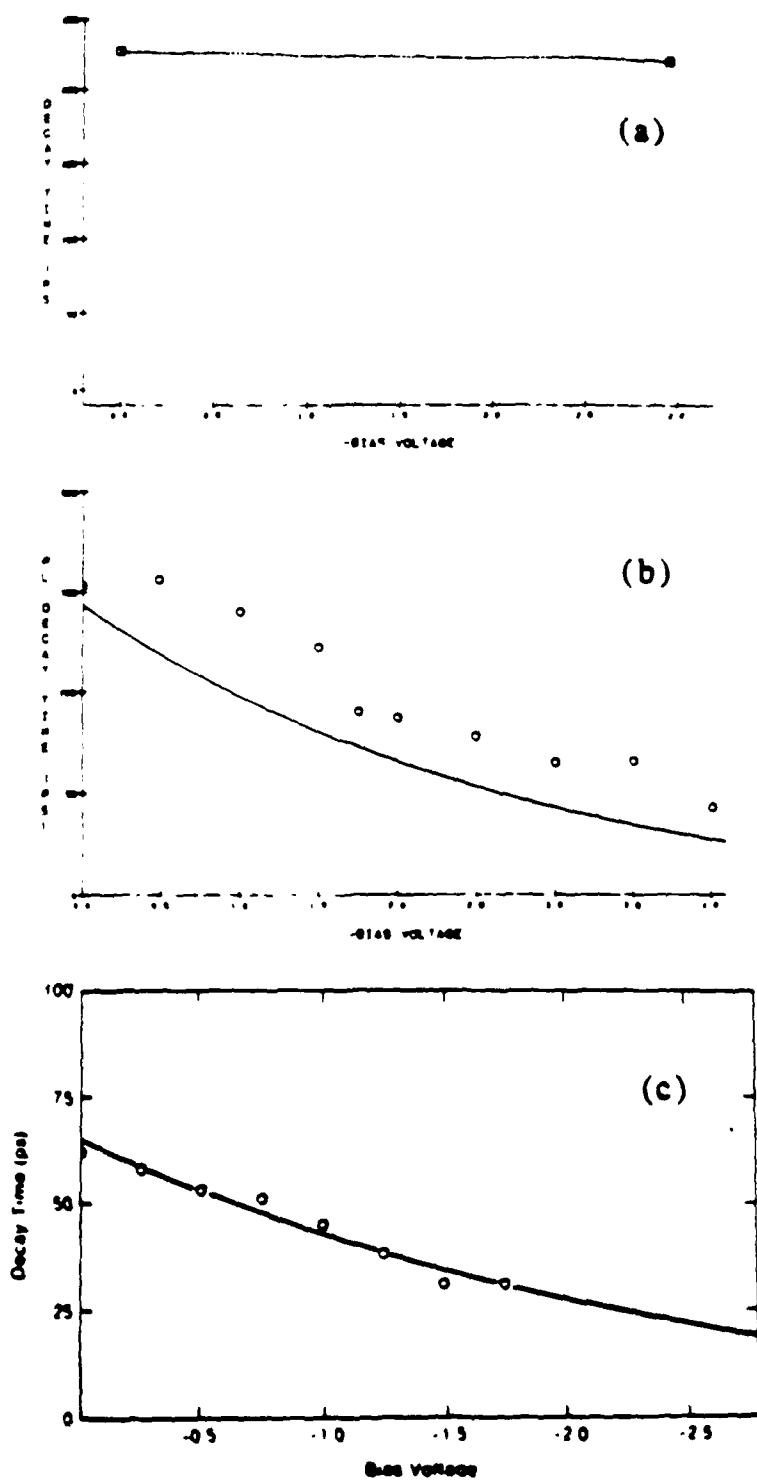


Figure 47: Luminescence decay time vs applied bias for different barrier heights. The barrier width  $b = 86\text{\AA}$  ( $85\text{\AA}$  for (c)). The Al compositions  $x$  are (a) 50%, (b) 38%, and (c) 30%. The dotted line of (a) is simply an aid to the eye; the solid lines of (b) and (c) are from the tunneling-out theory of section II.C.3.

The decay time  $\tau$  at zero bias is strongly dependent on the barrier width (as reported in Ref. 46) and on the height, as expected. Near zero bias, the decrease of  $\tau$  with applied bias is weak for samples with large barriers; for the data of Figs. 46(a) and 47(a) the recombination rate dominates at low bias. For higher fields the decay time decreases rapidly with applied field as the tunneling rate begins to dominate, except for 47(a), which evidently has such a high barrier that no tunneling is possible. In Fig. 46 it is apparent that the rate of decrease in tunneling time with applied field is higher for samples with wider barriers. The reason for this is that as the field increases to tilt the barrier, the effective barrier width to the tunneling electron becomes close for different samples.

### III.A.1.5. PL Stark Shifts

We mentioned previously that the time-resolved PL spectroscopy system allows the transition energy of the PL to be simultaneously determined. The resulting Stark shifts are shown in Fig. 48. For the sample with  $x = 50\%$ ,  $b = 86\text{\AA}$ , which shows no evidence of tunneling, the start shift is to lower energy by a few milli-electron-volts. This red shift is typical of previously studied QW structures<sup>71-73</sup>. However, in all other samples the Stark shift appeared as a small blue shift (peaking at about 3 meV for reverse biases in the range 1-2V). To our knowledge this is the first observation of a blue shift in the PL from a QW with an electric field. We discuss the possible origin of this shift in the following section.

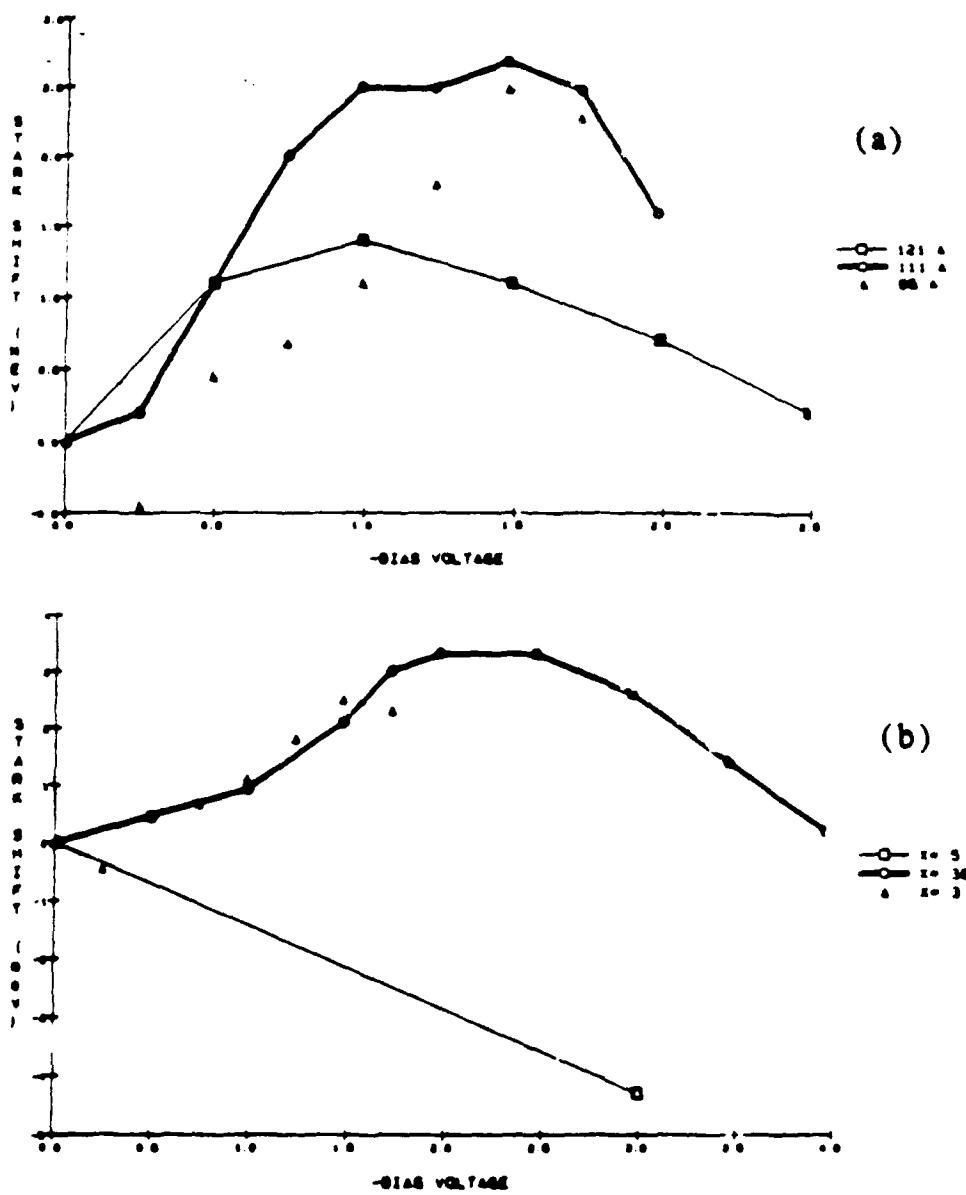


Figure 48: Stark shifts of the luminescence lines (a) for samples with different barrier width, and (b), for samples with different height (i.e., Al composition x). The error is about  $\pm 1$  meV.

### III.B. THEORETICAL INTERPRETATION

#### III.B.1. Electric Field vs Bias Voltage

Of course, before any comparison of experiment and theory is possible, it is necessary to determine the relation between the applied bias voltage and the magnitude of the electric field at the

location of the QW structure. Sze<sup>74</sup> has calculated the electric field inside a p-i-n diode. The difficulty with applying this calculation to a real MBE-grown AlGaAs p-i-n structure is that the "intrinsic" region, in fact, has a nonzero background p-type doping.<sup>75</sup> Typically this doping will be in the range  $10^{15}$ - $10^{16}$  cm<sup>-3</sup>, subject to the precise MBE growth conditions. This background doping will have a strong effect on the field profile inside the diode. Due to the difficulty of determining the doping level in each region of the diode independently, it is not clear how to accurately and self-consistently determine the field profile.<sup>76</sup> To our knowledge only Miller *et al.*,<sup>77</sup> have attempted to determine the field profile; however, they have not published the method they used to determine the (nonuniform) field.

We have taken the simplest approach, where we determine the depletion region width  $w$  by a capacitance-voltage (C-V) measurement,<sup>76</sup> and express the field simply as  $F=V/w$  where  $V$  is the external applied bias. (We find  $w$  is in the range of 4000 Å, so for example the field at -2V bias is about  $5 \times 10^4$  V/cm). Of course this neglects the built-in potential of the diode structure. However, as is shown in Miller *et al.*,<sup>77</sup> the background p doping tends to reduce the field in the region near the p-i junction, which is where the QW is situated in our samples. Hence our simple approach is a reasonable first approximation to the field. The good agreement we find between theory and experiment for the tunneling rates provides further confirmation that our approach is reasonable.

### III.B.2 Tunneling Time at Zero Field

In a semiclassical model, the tunneling rate  $1/\tau_T$  can be expressed as the product  $vT$  of the frequency  $v$  of the electron collisions with the barrier and the transmission probability  $T$  through the barrier. The oscillation frequency of the electron in the well under flat band conditions is

$$v = \frac{1}{d} \sqrt{\frac{E}{2m_w}} \quad (11)$$

where  $E$  is the electron energy (relative to the band edge),  $d$  is the well width, and  $m_w$  is the electron effective mass inside the well. For our structure,  $v$  is  $1.3 \times 10^{14}$  s<sup>-1</sup>.

The transmission coefficient  $T$  of the barrier is calculated in the usual way by matching the wavefunctions and probability current across the barrier. The only refinement necessary is that because the electron effective mass is different in the barrier from that in the well and continuum regions, continuity of probability current requires that  $(1/m^*)(d\psi/dz)$  instead of just  $d\psi/dz$  be continuous across the boundaries. The expression for  $T$  under flat band conditions is then

$$T = \frac{4r(V-E)E}{[(r-1)E+V]^2 \sinh^2 \left( \frac{b}{\hbar} \sqrt{2m_B(V-E)} \right) + 4r(V-E)E} \quad (12)$$

where  $r=m_B/m_w$ , and  $m_B$  ( $m_w$ ) is the effective mass in the barrier (well). The calculated tunneling times for the five samples are shown in Table 2, where we have assumed that 65% of the total band offset is the conduction band.

x = 0.3		b = 86 Å	
b (Å)	$\tau_T$ (ps)	x	$\tau_T$ (ps)
85	17	0.38	143
111	277	0.5	3400
121	809		

Table 2: Calculated electron tunneling times for the samples with no applied electric field.

### III.B.3. Field Dependence of the Tunneling Time

When an electric field is applied to the tunneling structure, both the electron oscillation frequency  $\nu$  and the transmission coefficient  $T$  will change. For the electric fields we are concerned with,  $\nu$  is essentially constant. It is easy to show that the correction term to the electron

oscillation period is  $(Fd/E)^2/32$ , which shows that even at a field of  $10^5$  V/cm the oscillation period changes only by 0.2%.

The transmission coefficient, however, has a very strong dependence on the field. Landau and Lifshitz<sup>45</sup> show that the dependence of the transmission coefficient on the evanescent wavevector  $\kappa$  of a particle inside a potential barrier is

$$T = \exp \left( -2 \int_0^b \kappa dz \right) \quad (13)$$

Thus, we may account for the tilting of the barrier by the electric field by writing the tunneling time

$$\tau_T = c \exp \left( \frac{2}{\hbar} \int_0^b \sqrt{2m(V-E-Fz)} dz \right) \quad (14)$$

as where  $c$  is a constant. We make the approximation that  $c$  and  $E$  do not change appreciably with the electric fields considered here. The proportionality constant is obtained by calculating the tunneling time at zero bias as described above.

The results of this calculation are shown as the solid curves in Figs. 44 and 45. We did not plot the result in Fig 48(a), since for this high barrier, the tunneling time is always much greater than the recombination rate, consistent with experimental observations. For the curve of Fig. 44(c), we used a zero-bias tunneling time of 65 ps instead of the calculated value of 17 ps. This was because we want to show that the expression (7) properly displays the field dependence of the tunneling time, even if it is difficult to accurately calculate the absolute magnitude of the rate.

The difficulty of calculating the precise values of the tunneling time is due to the exponential dependence of the time on the effective mass and barrier width and height. Thus small uncertainties in the sample parameters will produce very large changes in the calculated tunneling time. One might consider fitting the data by varying the sample parameters, (e.g., the barrier

width) but there are too many degrees of freedom to produce a unique fit. Hence we have assumed that the nominal sample parameters are correct, and have adjusted only the conduction band offset ratio. To illustrate the strong dependence of the tunneling time on the band offset ratio, we note the following. Suppose the conduction band offset ratio is in the range  $62 \pm 3\%$ . Then the tunneling time for the sample of Fig. 45(a) will be 1.3, 2.1, and 3.2 ns for 59%, 62%, and 63%, respectively. The 59% value yields results that are inconsistent with experimental observations at high applied bias. Using a ratio of 65% produces a reasonable fit for all the data except for Fig. 44(c).

### III.B.4. Luminescence Blue Shift

Previously we noted that the Stark shift of the PL line appears as a blue shift for all structures which display tunneling. This is in contrast to the usually observed red shift. Thus the blue shift is intimately connected with the tunneling. We believe that considerable further experimental and theoretical work will be necessary to understand the cause of this shift. Nevertheless, in this section we will make a few remarks on its possible origin.

One possibility is that as the electrons tunnel from the well, a dipole field builds up that raises the potential barrier seen by the remaining electrons in the well. The higher barrier increases the confinement energy of the electrons in the well; hence the shift in PL to higher energy. This explanation turns out not to be reasonable, since the electrons that have tunneled out of the well will be swept away from the barrier region in about a picosecond. Thus a dipole field cannot build up. Furthermore, the possibility of a transient dipole field can be ruled out, since this would produce a dynamic Stark shift, contrary to observations.

Another possibility is that the shift is due to the holes that are left behind in the well. The holes are strongly confined in the well. Therefore when most of the photogenerated electrons tunnel from the well, the hole lifetime becomes very long since there are no electrons with which they may recombine. If the lifetime is comparable to or longer than the 10-ns time between pump

pulses, a substantial positive charge density will build up in the well. Thus the QW will behave in a manner very similar to a p-doped QW.

It is well known that free carriers screen the electron-hole coulomb interaction, thus reducing the exciton binding energy.<sup>78</sup> In fact Sanders and Chang<sup>79</sup> have calculated the exciton binding energy as a function of dopant charge density for p-type modulation-doped wells. Such a reduction in binding energy would, of course, produce a blue shift in PL energy. However, a competing effect is band-gap renormalization, which shifts the band edge to lower energy with increasing density.<sup>80</sup> There have been some calculations of the optical properties of doped QW's, but they have principally concentrated on the absorption spectrum.<sup>79,81,82</sup> Obviously, further theoretical and experimental study is required to determine if the blue shift is, in fact, a many-body effect.

### III.C. TUNNELING FROM ASYMMETRIC QUANTUM WELLS

#### III.C.1. Photoluminescence Experiments

In this section, we present the results of time-resolved and continuous (cw) photoluminescence (PL) spectroscopy experiments we performed to observe tunneling in asymmetric DQW structures. As we mentioned in an earlier section, the time resolution of the streak-camera PL system is about 20 ps; hence, one cannot observe coherent resonant tunneling with this system. Nevertheless, many interesting tunneling-rate processes are observable.

The DQW structures used in this study are of an entirely novel design. The two QW's are of different widths, but the Al composition of the wider well is adjusted so that under flat band conditions the electron energy levels are near resonance, and the hole energies are sufficiently different so that the PL energies of the two wells are well separated.<sup>83</sup> This is possible because the heavy hole mass is much larger than the electron mass, so the holes sit close to the top of the valence band in each QW. As we will discuss below, this makes possible at reasonable electric fields the existence of a "charge-transfer" (CT) state, in which the electrons and holes are in

different wells. In these experiments, we have directly observed for the first time the buildup of a CT state via electron and hole tunneling in opposite directions in asymmetric DQW structures.

### III.C.1.1. Sample Design

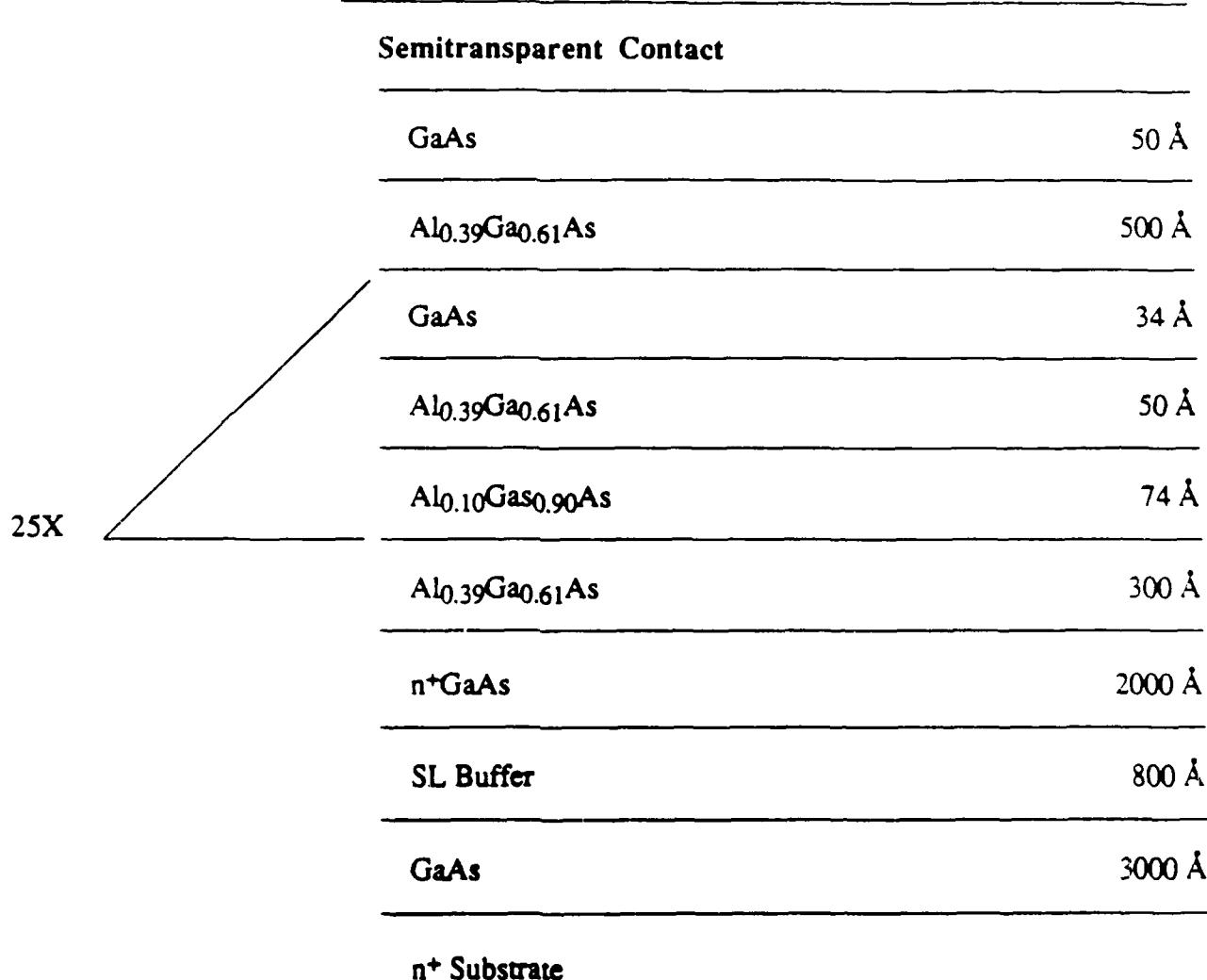


Figure 49 Double QW sample structure for PL studies. The nominal growth parameters in angstroms are given. For sample A, the barrier width  $b$  is 50 Å; for sample B,  $b=100\text{Å}$ .

The samples used for this study were grown by molecular beam epitaxy at the central research laboratory of Thomson-CSF in Orsay, France. The growth layers are shown schematically in Fig. 49.

The nominal growth parameters are as follows. A 74-Å  $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$  QW (QW1) is coupled to a 34-Å GaAs QW (QW2) through an  $\text{Al}_{0.45}\text{Ga}_{0.55}\text{As}$  barrier. Two samples were grown from this study; sample A has a thin (50-Å) barrier so the electron states are somewhat delocalized over the two wells, and sample B has a thick (100-Å) barrier so the electron states are strongly localized in each QW. Twenty-five periods of the DQW structure were grown on an  $n^+$  GaAs substrate and  $n^+$  GaAs buffer layer. The doping level is  $5 \times 10^{17} \text{ cm}^{-3}$ . (A 20-period  $n^+$  AlAs/GaAs superlattice was used to smooth the substrate surface.) A semitransparent Al, Schottky contact was evaporated on the top surface so that the effect of an electric field applied along the growth direction could be studied.

### III.C.1.2. Electronic States

The electron and hole states for the above-described DQW system have been calculated by B. Vinter.<sup>84</sup> He obtained the exact wavefunctions and energies by numerically integrating the Schrödinger equation for the coupled well system, rather than starting with the localized states as a basis. The band diagram and results for the wavefunctions for sample A are shown in Fig. 50, both under flat band conditions and with an applied electric field (reverse bias). The band offset ratio was assumed to be 0.62. Under flat band conditions the hole states are strongly localized in each well (due to their heavy mass), but the electrons are somewhat delocalized as expected. With a moderate applied electric field, the electron states become localized in each QW. For sample B, the electrons are always strongly localized.

The energy levels vs applied field shows strong level repulsion near resonance due to the coupling of the wells. Unfortunately, the PL transition energies predicted by this calculation do not agree very well with the observed values (7038 Å for QW1 and 7160 Å for QW2 at 6 degrees K). Therefore, we have had to revise the sample growth parameters in order to fit the observed PL

spectra. If we assume that the AlAs growth rate  $R_{Al}$  was as desired, then the PL energies can be fit by varying the GaAs growth rate  $R_{Ga}$ . The thickness of each AlGaAs layer is then given by  $(R_{Al}+R_{Ga})t$ , where  $t$  is the growth time for that layer. The Al composition  $x$  is given by  $x=R_{Al}/(R_{Al}+R_{Ga})$ .

By varying the GaAs growth rate, we determined values of the QW widths and Al compositions that gave a reasonable fit to the PL energies. The results were that QW2 is only 26 Å wide, the barrier is 43 Å (86 Å) for sample A (B) with  $x=0.45$ , and QW1 is 50 Å wide with  $x=0.15$ .

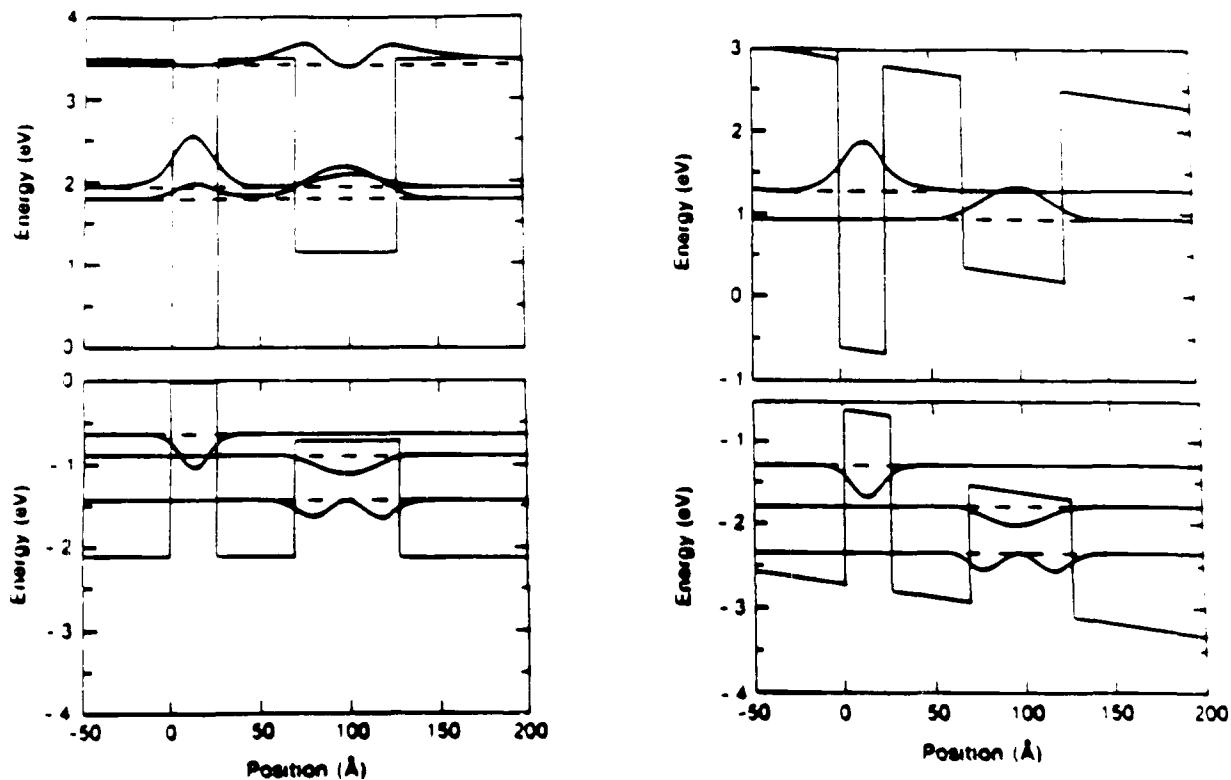


Figure 50. Band diagram and calculated electron and hole states for sample A (50-Å barrier), (a) with no electric field, and (b) with an electric field of 30 kV/cm. Similar results are obtained for sample B, with the important difference that the electron states are strongly localized in each QW.

### III.C.1.3. Experimental Setup

The samples were held in a cryostat at a temperature of 6 K. The laser and time-resolved PL setup have been described in detail previously. The laser dye used was Pyridine 1; the laser was tuned with a single-plate birefringent filter (Lyot) and an uncoated 5- $\mu$ m pellicle to 6947  $\text{\AA}$ , so electron-hole pairs were generated in both wells at  $t = 0$ . For the time-resolved experiments discussed here, we estimate that each laser pulse injected a pair density of approximately  $10^{11} \text{ cm}^{-2}$  in each well. (This results in a band-filling in the conduction band of 3 meV and in the valence band of 0.4 meV.) The resulting PL was dispersed by a 0.32 m monochromator with 300 1/mm grating across the entrance slit of the synchroscan streak camera. The streaked image was integrated on a 2-D SIT detector, allowing us to obtain PL spectra with a spectral resolution of about 3 meV and a temporal resolution of 22 ps (i.e., the FWHM of the dye laser output).

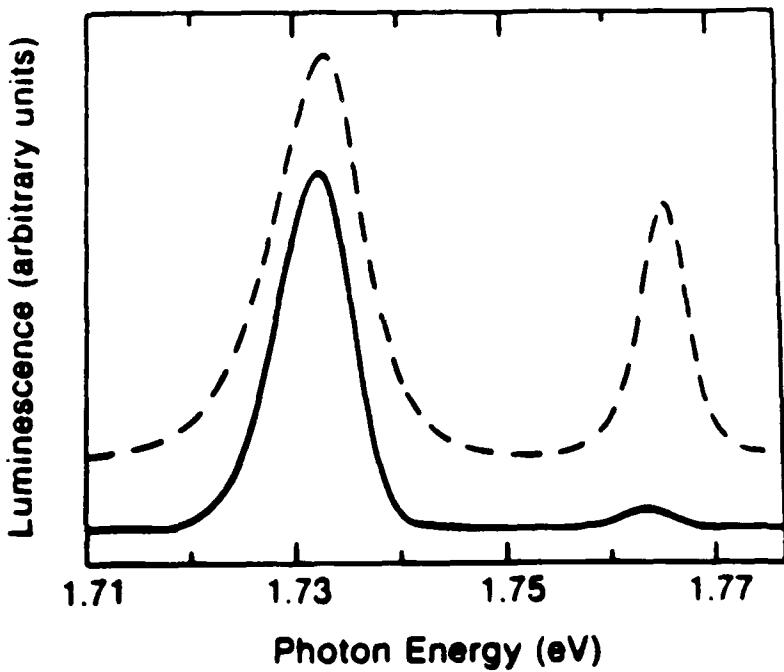


Figure 51. Time-integrated (cw) PL spectra for samples A and B with zero applied bias at a temperature of 6 K.

The time-integrated PL spectra presented here were obtained at Thomson-CSF using a separate system with sub-milli-electron volt spectral resolution. CW PL spectra are shown for

both samples with zero bias voltage in Fig. 51. For sample B, the  $\omega_1$  and  $\omega_2$  transitions are about the same magnitude; but for sample A, the  $\omega_1$  intensity is much smaller than for  $\omega_2$ , indicating stronger tunneling processes for this sample as will be discussed below.

### III.C.1.4. Experimental Results

Time-resolved spectra for sample B (thick barrier) are shown in Figs. 52 and 53 for reverse bias voltages from 0 to -6V. The spectra show the scattered pump light (defining the time origin) and two PL lines corresponding to recombination within each of the two wells. (The small peaks just after the pump pulse at approximately the same wavelength are due to an unavoidable spurious reflection of the laser onto the streak camera and should be ignored. Also it should be noted that the absolute intensities cannot be compared for different plots.) The high energy PL line corresponds to the QW1 transition, and the lower energy line to QW2. The PL rise and decay times change very little with applied bias, as is shown in Fig. 54. The Stark shift of the PL peaks, plotted in Fig. 55 is observed to be a few meV to the red as expected.<sup>85</sup>

In the cw PL spectra shown in Fig. 56, however, we observe a third PL peak which shifts strongly and approximately linearly towards the red with applied bias. In the narrow barrier sample (A) this peak appears even more clearly in the cw spectra (Fig. 57), and presents a stronger Stark shift. The origin of this peak can be elucidated by examining the time-resolved PL spectra for the thin barrier sample; the line comes from radiative recombination between electrons in QW1 and holes in QW2, which we refer to as the charge-transfer (CT) state.

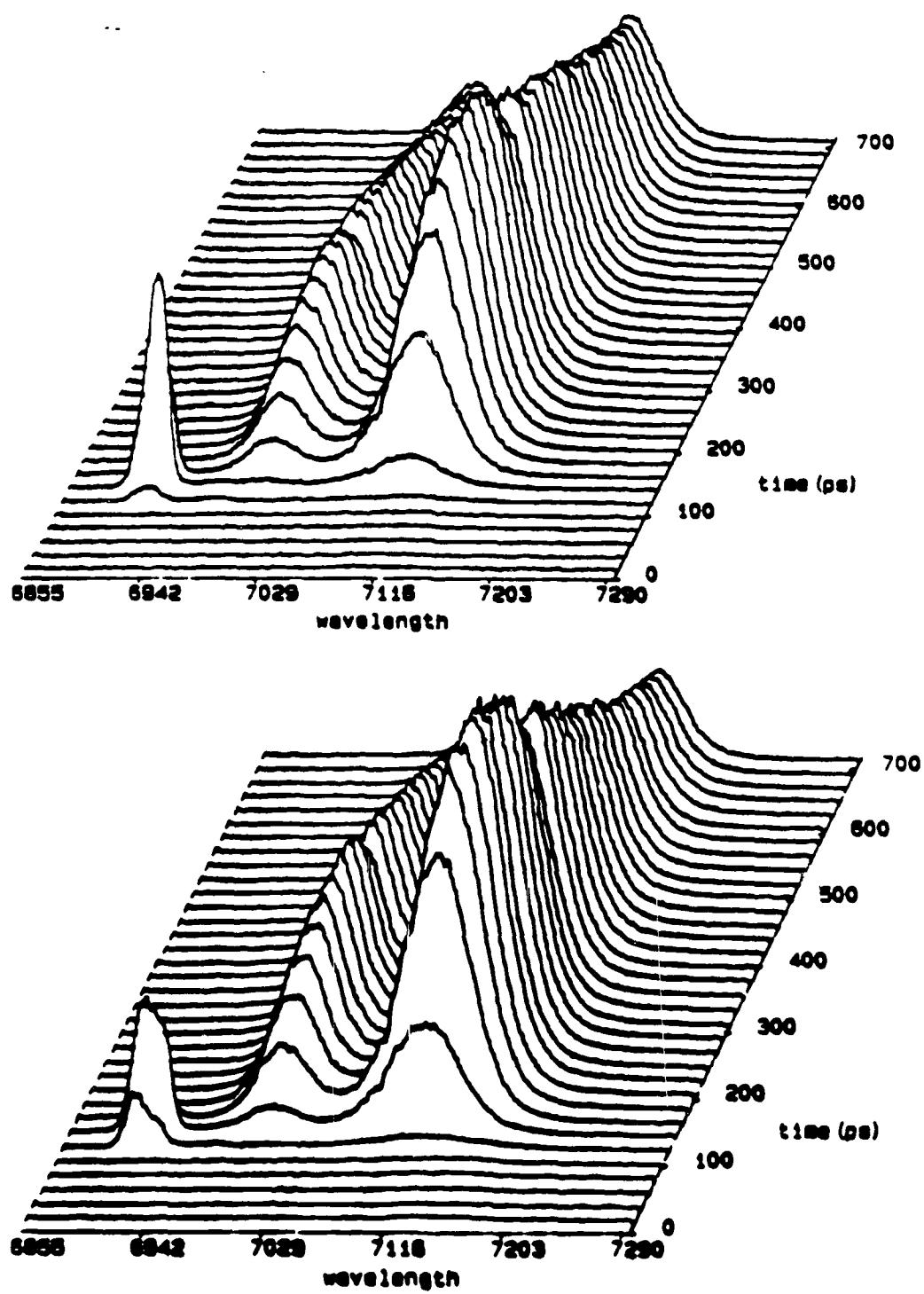


Figure 52: Time-resolved PL spectra of sample B for (a) zero applied bias and (b) -2V reverse bias.

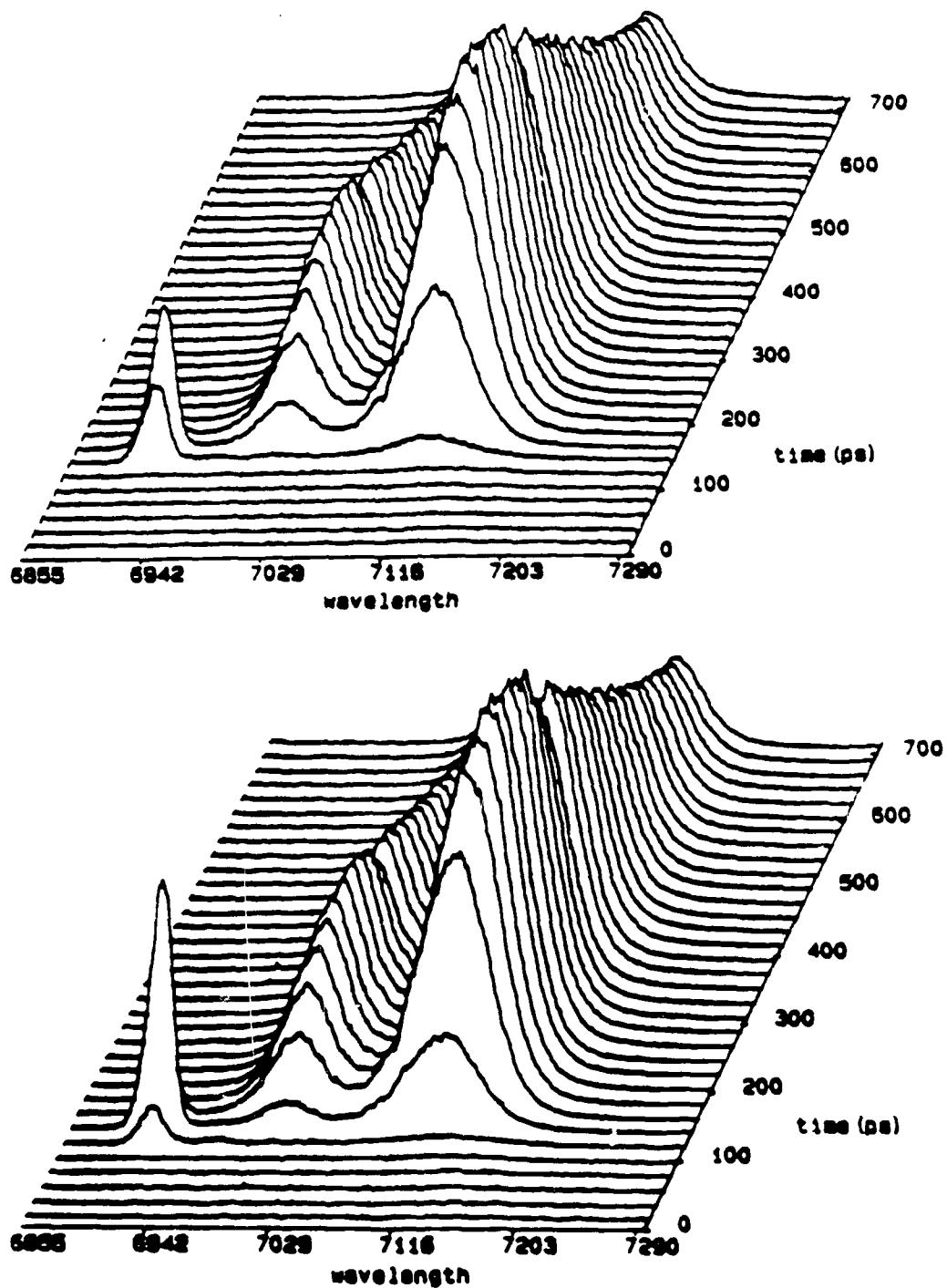


Figure 53: Time-resolved PL spectra of sample B for (a) -4V and (b) -6V bias.

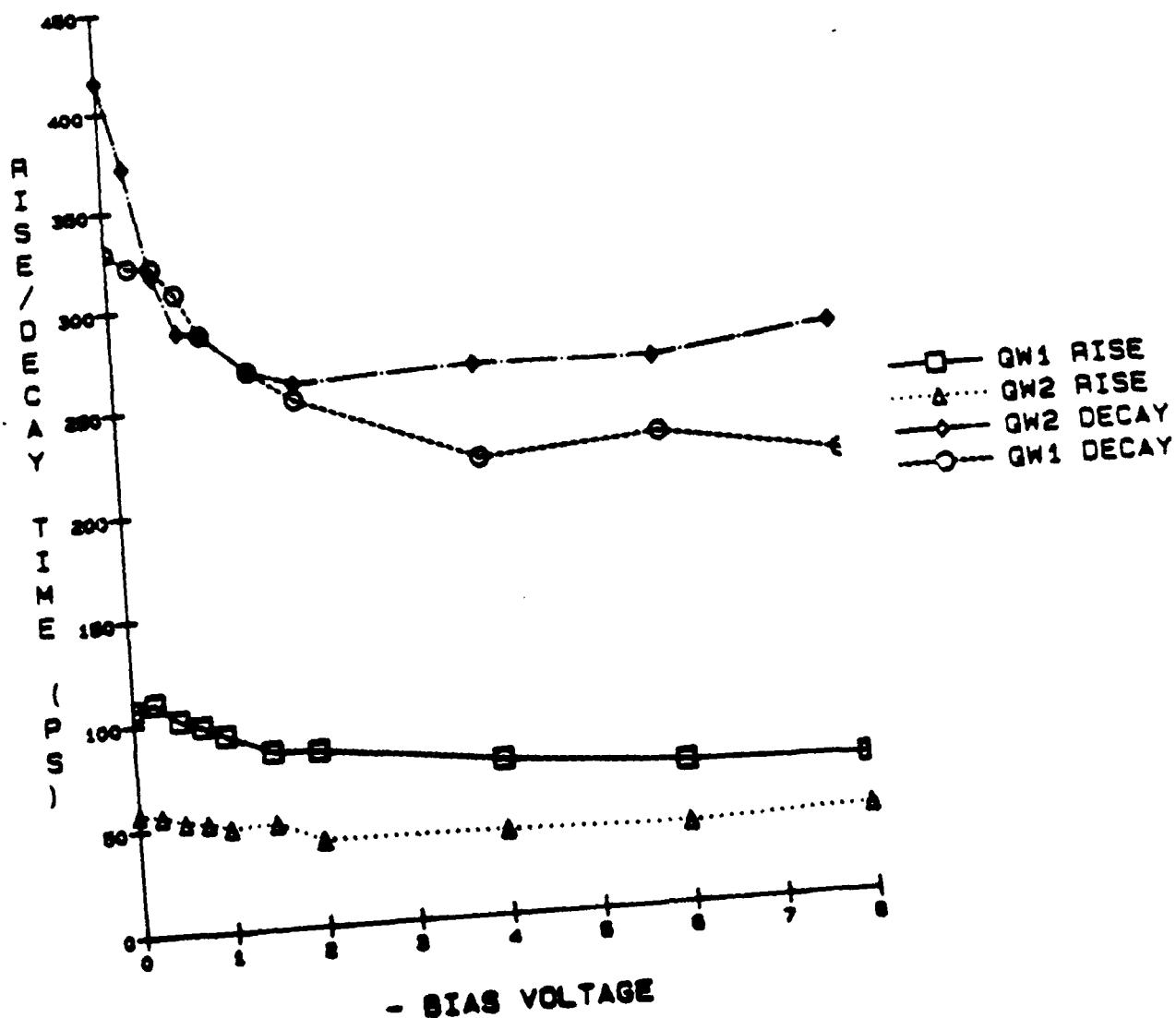


Figure 54: Rise times (10-90%) and decay times (1/e single exponential fit) for the two PL lines of sample B.

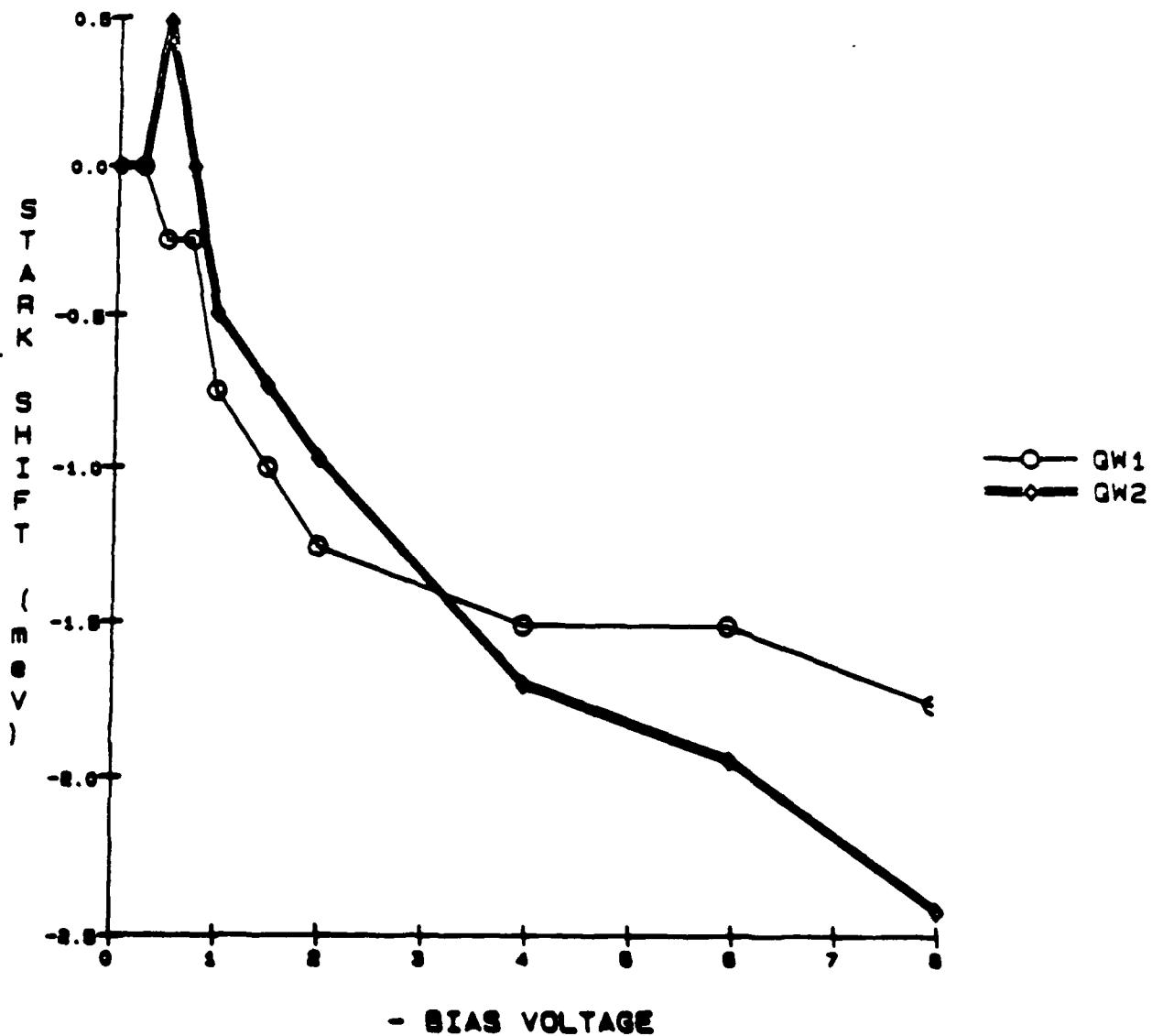


Figure 55: Stark shifts of the two PL lines of sample B.

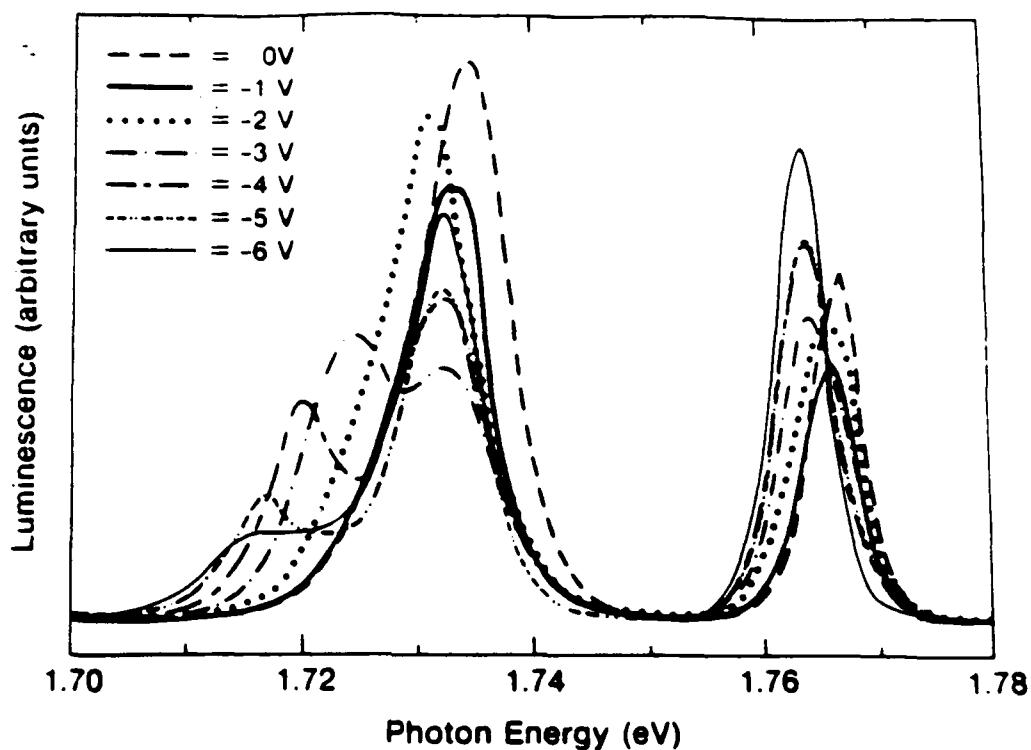


Figure 56: CW PL spectra of sample B at different applied voltages.

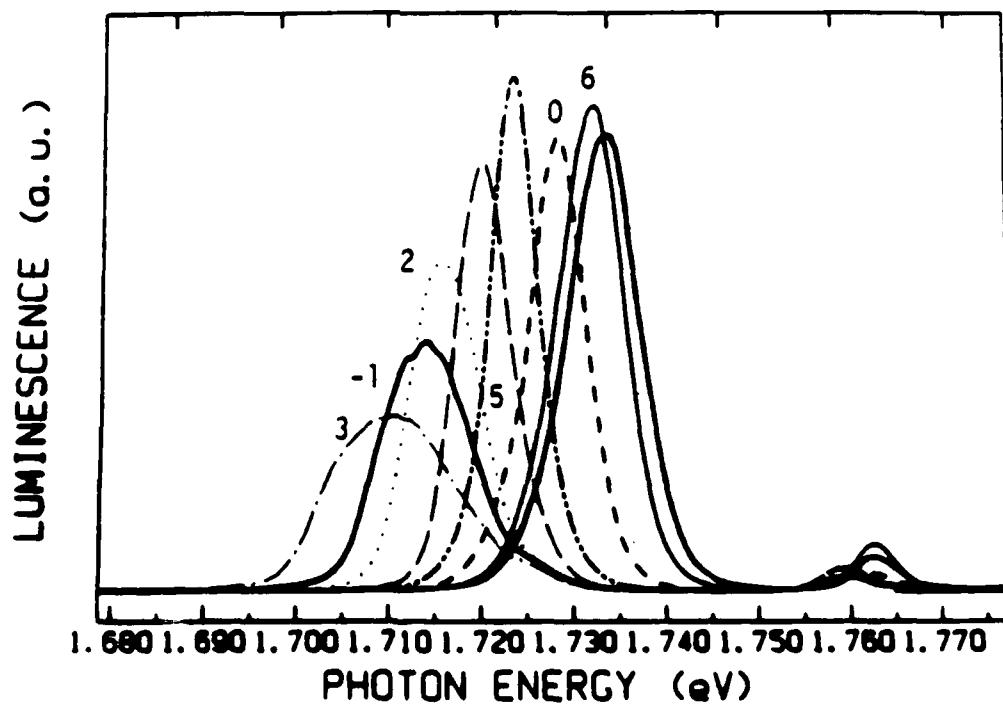


Figure 57: CW PL spectra of sample at different applied voltages.

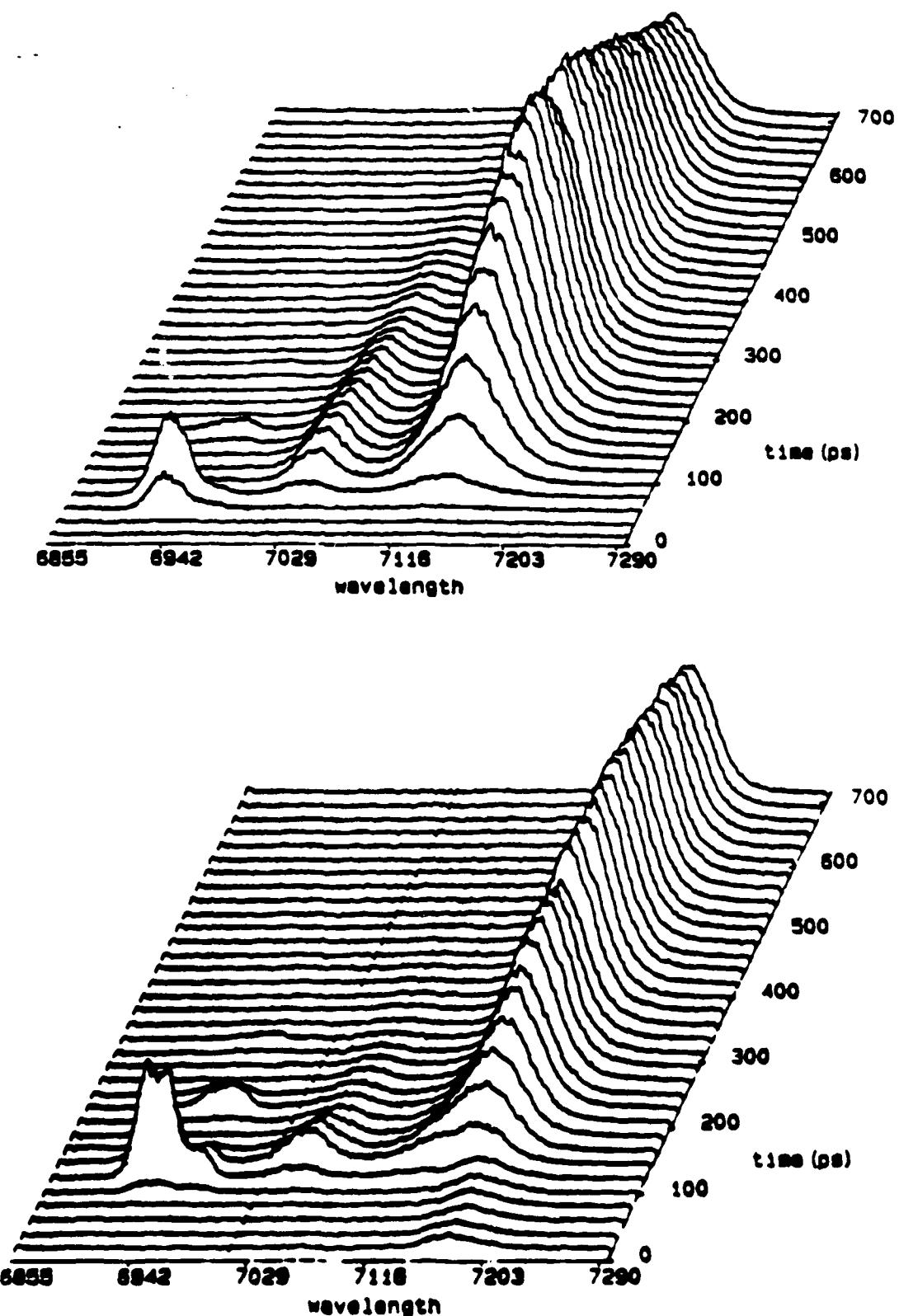


Figure 58: Time-resolved PL spectra of sample A for (a) zero applied bias, and (b) -2V reverse bias.

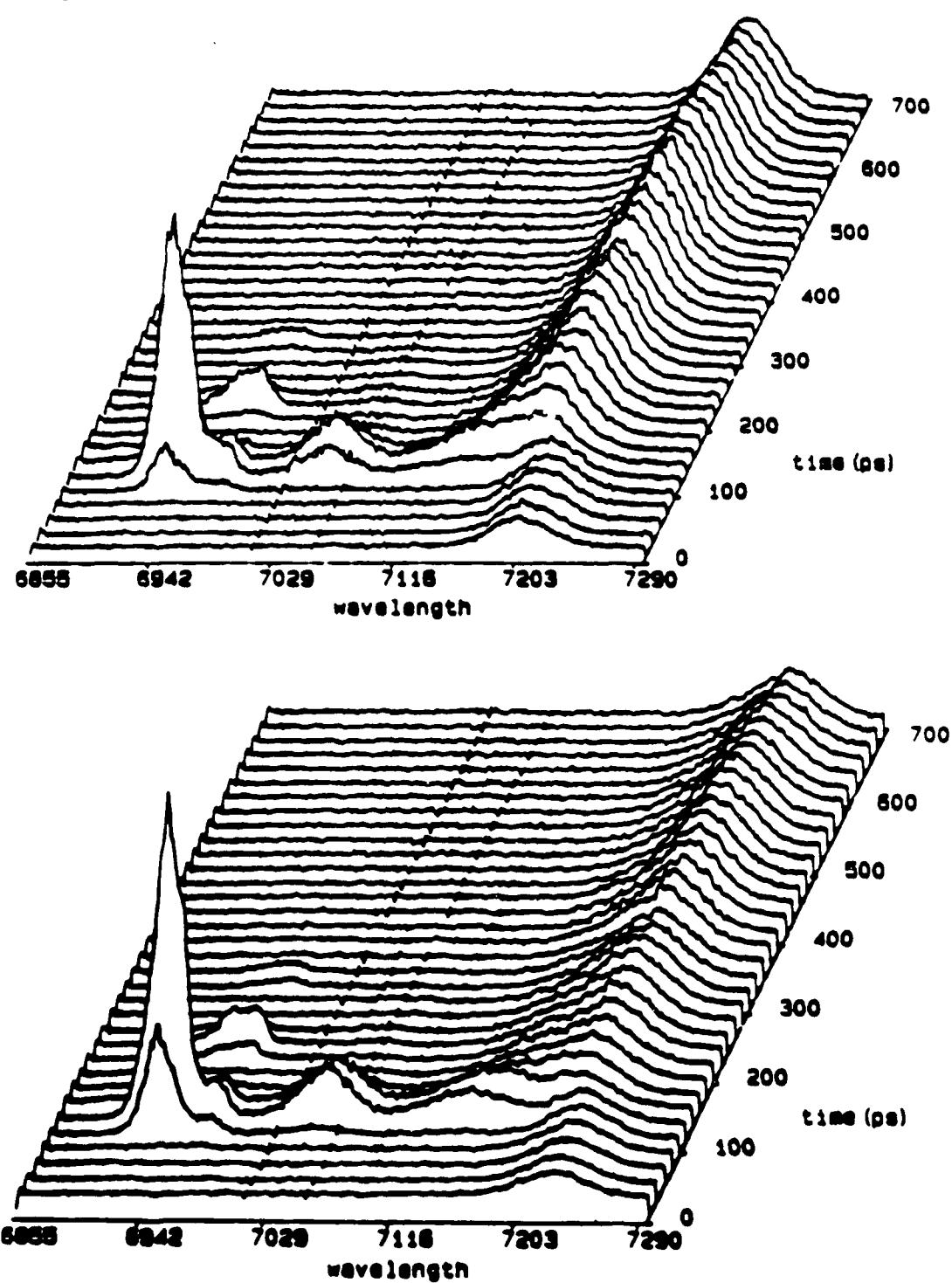


Figure 59: Time-resolved PL spectra of sample A for (a) -3V and (b) -4V

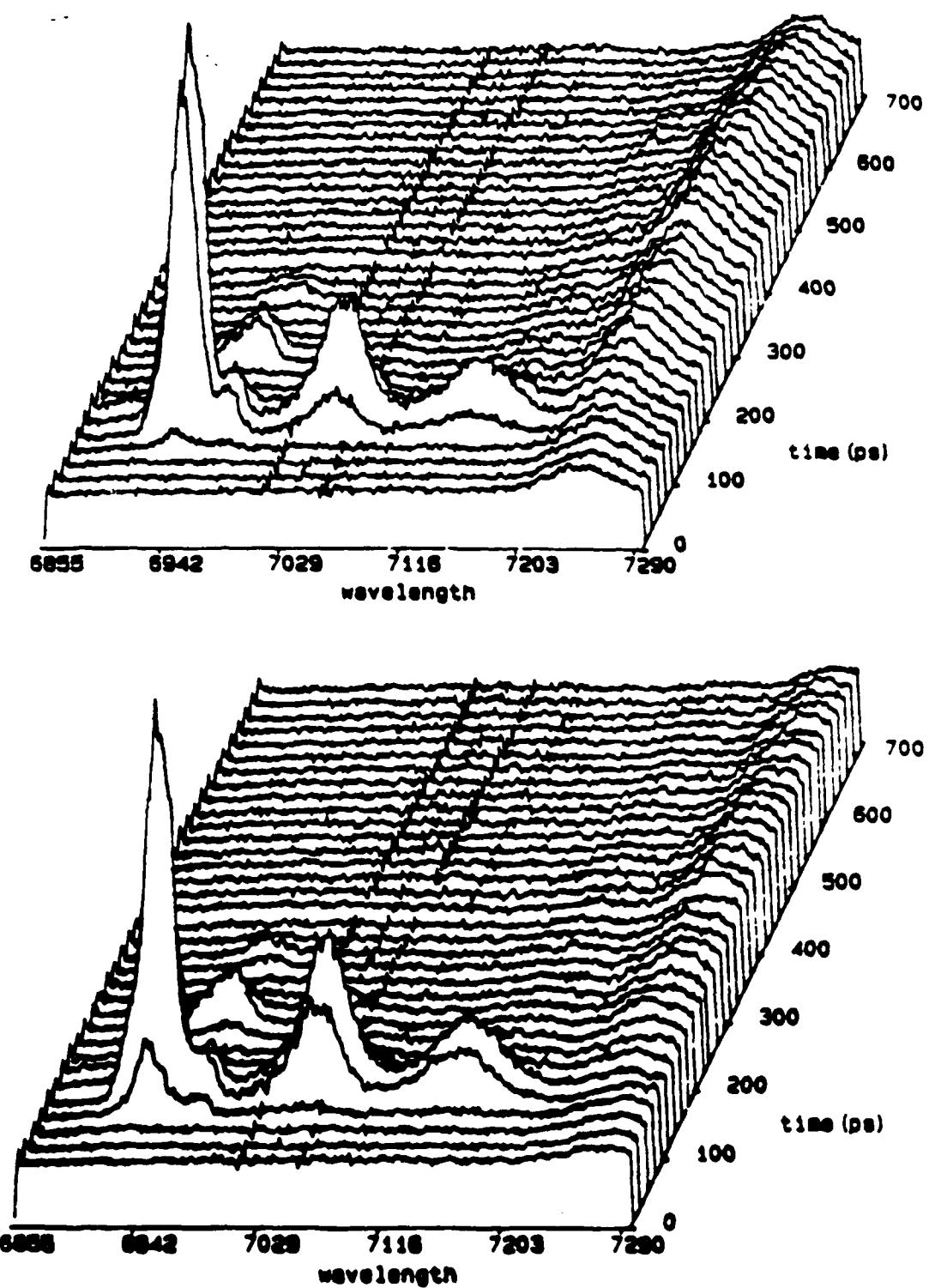


Figure 60: Time-resolved PL spectra of sample A for (a)  $-0.5V$  and (b)  $-0.6V$

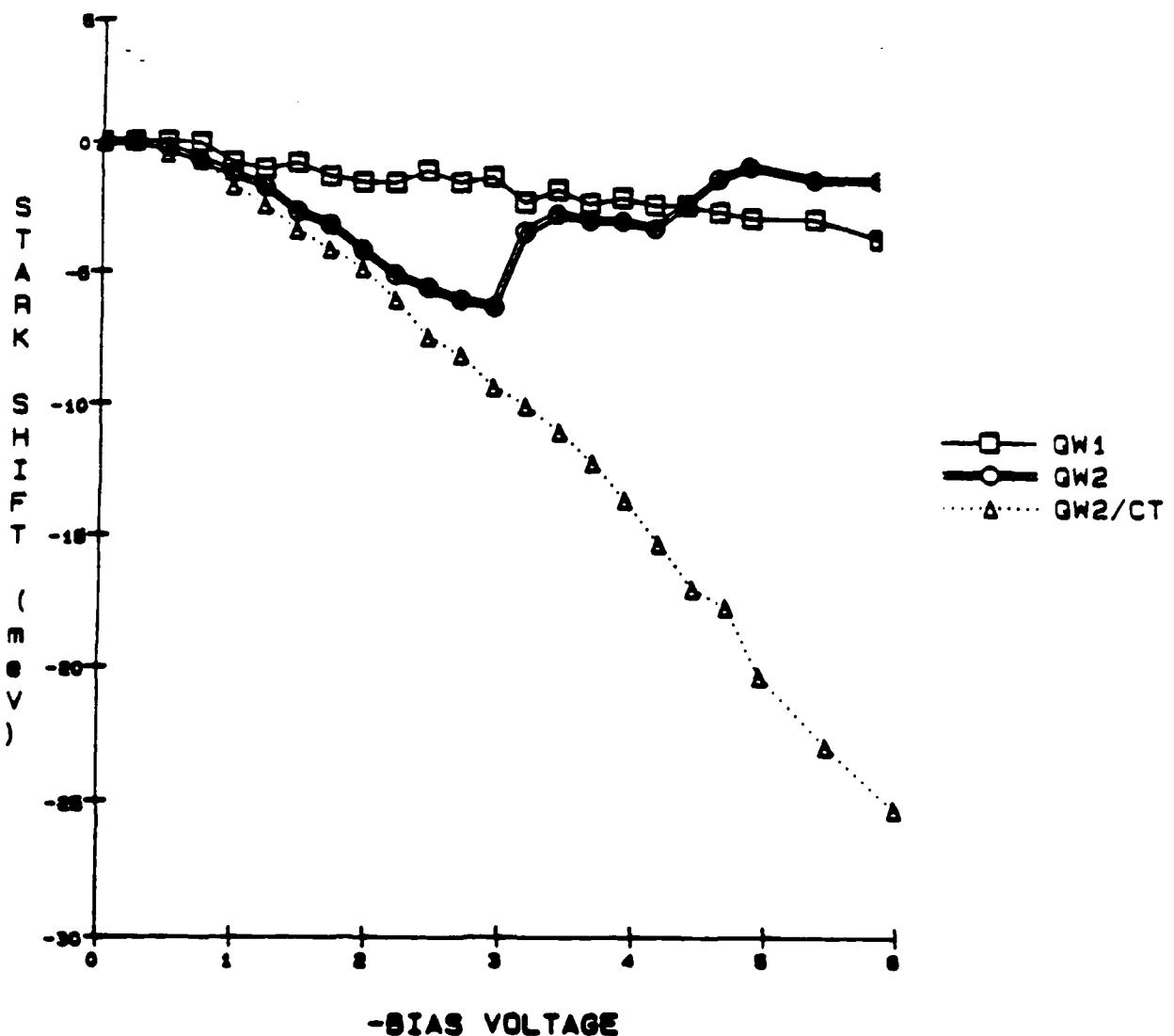


Figure 61: Stark shifts of the three lines observed in the time-resolved PL spectra of sample A.

A selection of the time-resolved PL spectra vs applied bias is shown in Figs. 58-62 for sample A. At low bias, one sees the scattered pump light and two PL lines corresponding to the  $\omega_1$  and  $\omega_1 + \omega_{CT}$  transitions of Fig. 50. It should be noted that at zero bias, the PL of QW2 and the CT state have the same transition energy. As the electric field is increased, the  $\omega_1$  and  $\omega_{CT}$  lines separate and  $\omega_{CT}$  is strongly Stark-shifted to lower energy. This PL has a long lifetime that, in fact, approaches or exceeds the 10-ns time interval between pump pulses and the synchroscan

sweep cycle time; hence the PL signal that appears roughly constant vs time on the streaked spectrum. (This is clearly seen as a signal at "t < 0".)

The Stark shifts of the three PL lines vs bias are shown in Fig. 61. For bias voltages less than about -3V, the  $\omega_2$  and  $\omega_{CT}$  lines are not well separated. The  $\omega_2$  wavelength was determined by the peak position near  $t = 0$ , and the  $\omega_{CT}$  wavelength by the peak position at "t < 0" except where this component did not appear, in which case the position at  $t \sim 600$  ps was used. The  $\omega_1$  and  $\omega_2$  lines show a shift of a few milli-electron-volts typical of an exciton confined in a QW. The Stark shift of the  $\omega_{CT}$  line, however, is much stronger, and is approximately linear with applied bias. The strong red shift and long decay time indicate that this PL, in fact, is the result of recombination of electrons and holes in the CT state. The CT state is built up by electrons tunneling from QW2 to QW1 and holes tunneling from QW1 and QW2. The long lifetime is the result of the small overlap of the electron and hole wavefunctions, but the CT luminescence is nevertheless observable due to the nearly complete charge separation that takes place. This explains why the CT luminescence is observable in the cw PL spectra but not in the time-resolved spectra for sample B. In this case, the charge separation is not so strong due to the much slower electron and hole tunneling rates between the wells, so although the long-lived radiative recombination is observed as a line in the cw PL spectrum, the instantaneous intensity is low and, therefore, extremely weak on the time-resolved spectrum. In Fig. 62 the time-resolved spectrum of sample B at -6V bias is shown for a higher pump intensity than that for Fig. 53, and one can observe just above the noise a very weak red-shifted component at several hundred picosecond delay that is from the CT transition. (It is likely observable in this case since a larger charge separation apparently occurs with larger band-filling.)

One further observation confirms that the origin of the strongly red-shifted PL line is from the CT state. The Stark shift of this line is dynamic, since as the carriers recombine the screening of the applied electric field by the CT state is reduced and the red shift increases. This is most easily observed by comparing the CT wavelength just after the pump pulse with the CT PL wavelength at a very long delay time (which in practice is done by observing the PL at "t < 0" on

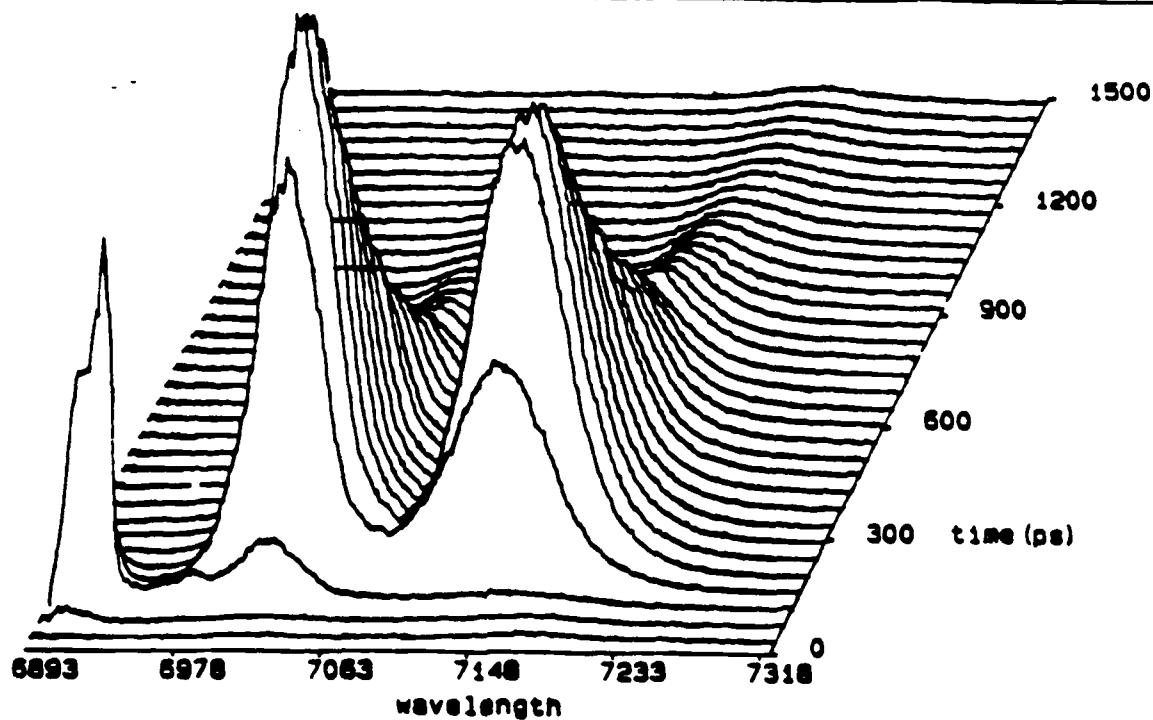


Figure 62: Time-resolved PL spectrum of sample B at -6V bias.

the synchroscan streak image). This is only reliable in the range 2-4 V, where the CT lifetime is not so long that the principal part of the signal at  $t = 0$  is, in fact, due to PL emitted at  $t \gg 0$ .

We mentioned previously that the CT state can have a lifetime that can exceed the pump pulse period. As a consequence, a dc charge can build up in wells (this is most dramatically apparent in Fig. 60, where all the CT luminescence signal comes from these steady-state separated electron-hole pairs). This charge screens the applied field, thus reducing the observed CT Stark shift. Since the CT Stark shift effectively gives the average electric field in the DQW region, the separated charge density can be estimated by comparing the observed PL Stark shift to the calculated electric field in the Schottky diode without the presence of charge carriers. We can then estimate an effective lifetime  $\tau$  for the separated carriers from  $n = g\tau$  where  $g$  is the generation rate.

This procedure is accurate for cw experiments and yields good approximate results for the time-resolved experiments since the lifetime  $\tau$  is comparable to or longer than the pump pulse period of 10 ns. We make the simplifying assumption that complete charge separation takes place after each pump pulse; this is reasonable for sample A, particularly at voltages greater than about

-3V, but is not reasonable for sample B. Since the separation between the QW's is much less than the laser pump diameter, to a very good approximation the electric field due to the separated charges may be calculated by assuming infinite sheets of charge with areal charge density  $\sigma$ . This screening field is then

$$E = \sigma / \epsilon \epsilon_0,$$

where  $\epsilon$  is the dielectric constant of GaAs. For example, a pair density of  $10^{11} \text{ cm}^{-2}$  will give a field of  $1.5 \times 10^4 \text{ V/cm}$ . The effective cw generation rate for the time-resolved experiments described above was  $10^{19} \text{ cm}^{-2}\text{s}^{-1}$ . The Schottky barrier height of the Al contact on the GaAs surface is 0.8 eV,<sup>86</sup> and the not-intentionally-doped region of sample A is nominally 9710 Å; thus the built in electric field in the DQW region is approximately  $8.2 \times 10^3 \text{ V/cm}$ . To a good approximation (*i.e.*, ignoring barrier-lowering), the total field under reverse bias is just the sum of the built-in field and the applied field. The results for the carrier density and CT state lifetime for the time-resolved experiments on sample A are given in Table 3; also given is the band-filling due to the dc charge buildup for both electrons and holes.

Bias Voltage (V)	$n(\text{cm}^{-2})$	$\tau(\text{ns})$	$\Delta E_c(\text{meV})$	$\Delta E_v(\text{meV})$
-1	$5.7 \times 10^{10}$	5.7	1.7	0.3
-2	$1.1 \times 10^{11}$	11.0	3.3	0.6
-4	$1.9 \times 10^{11}$	19.0	5.8	1.0
-6	$2.5 \times 10^{11}$	25.0	7.6	1.3

Table 3: Separated charge densities, CT state lifetime, and conduction and valence band filling for sample A time-resolved experiments at various reverse bias voltages.

The lifetimes are surprisingly short. A separate experiment was performed to investigate the dependence of the CT state lifetime on the injected carrier density. Table 4 shows the results for the average separated charge densities  $n$  and CT lifetime  $\tau$  vs injected carrier density  $n$ , *i.e.*, the

electron-hole pair density injected by a single laser pulse). The bias voltage for this data was -6V, so the charge separation should be complete.

$n_i(\text{cm}^{-2})$	$n(\text{cm}^{-2})$	$\tau(\text{ns})$
$6.6 \times 10^{11}$	$2.4 \times 10^{11}$	3.6
$9.9 \times 10^{10}$	$1.8 \times 10^{11}$	18.0
$3.2 \times 10^{10}$	$1.6 \times 10^{11}$	50.0

Table 4: Average separated charge densities and CT state lifetime vs. injected charge density for sample A, -6V bias.

(There is a discrepancy between the results of this experiment and the results presented in Table 3 for the lifetime at an injected carrier density of  $10^{11} \text{ cm}^{-2}$ ; this is most likely due to error in determining the pump laser spot size precisely at the location of the cryostat in the sample, resulting in an error in calculating the injected carrier density.)

Apparently there is a strong dependence of the lifetime on the injected carrier density, but the average separated charge density depends only weakly on the initial injected density. This indicates that the observed short lifetime may be due to band-filling. This notion is supported by cw PL experiments performed at Thomson-CSF,<sup>88</sup> where a strong dependence of the lifetime on the pump intensity was also observed: for sample A at -3V bias and  $300 \text{ W/cm}^{+2}$  pump intensity,  $\tau$  is 20 ns; but at  $3 \text{ W/cm}^{+2}$ ,  $\tau$  is  $1.2 \mu\text{s}$ .

A more refined analysis treating the charged QW's in a self-consistent manner would be required to elucidate the processes that determine the lifetime  $\tau$ , which it must be noted is not necessarily the radiative lifetime for the CT transition.

The decay rates for sample A of the  $\omega_1$  and  $\omega_2$  lines vs. bias are also extremely difficult to interpret. The  $\omega_1$  decay rate vs. applied bias is plotted in Fig. 63. We have not plotted the  $\omega_2$  decay rate, since at low bias the situation is complicated by the fact that the  $\omega_2$  and  $\omega_{\text{CT}}$  lines are

very close in energy, and therefore are indistinguishable on the PL spectra (see Figs. 58 and 59). The difficulty of defining a decay rate for the  $\omega_2$  line except at high bias is clearly shown in Fig. 64, where we have plotted the data of Figs. 59(a) and 60(a) in a different way, so that the time dependence of the different spectral components may easily be seen. It is obvious that the measured decay time would depend entirely on the spectral window around  $\omega_2$  selected. At high

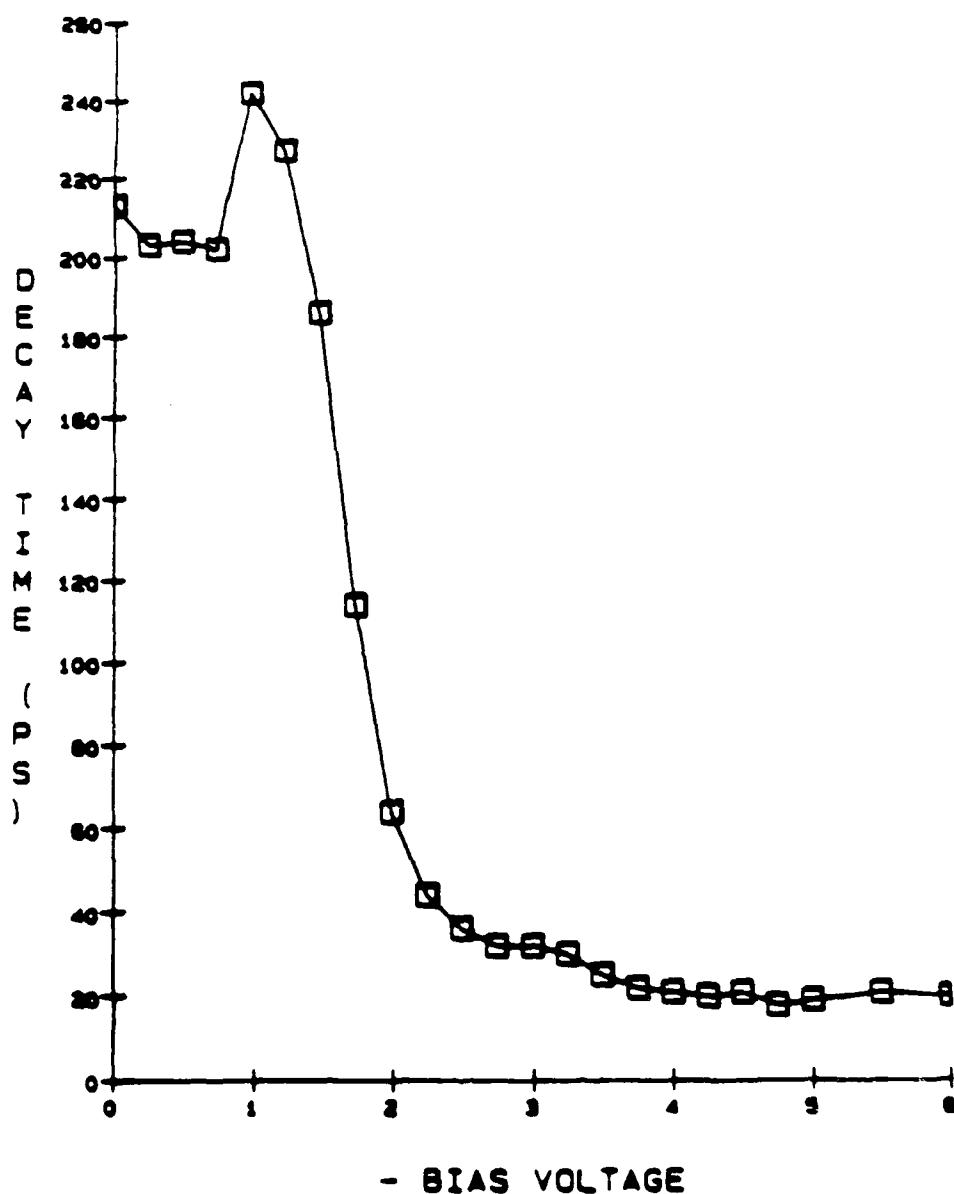


Figure 63: Decay time of the  $\omega_1$  PL line vs. applied bias for sample A.

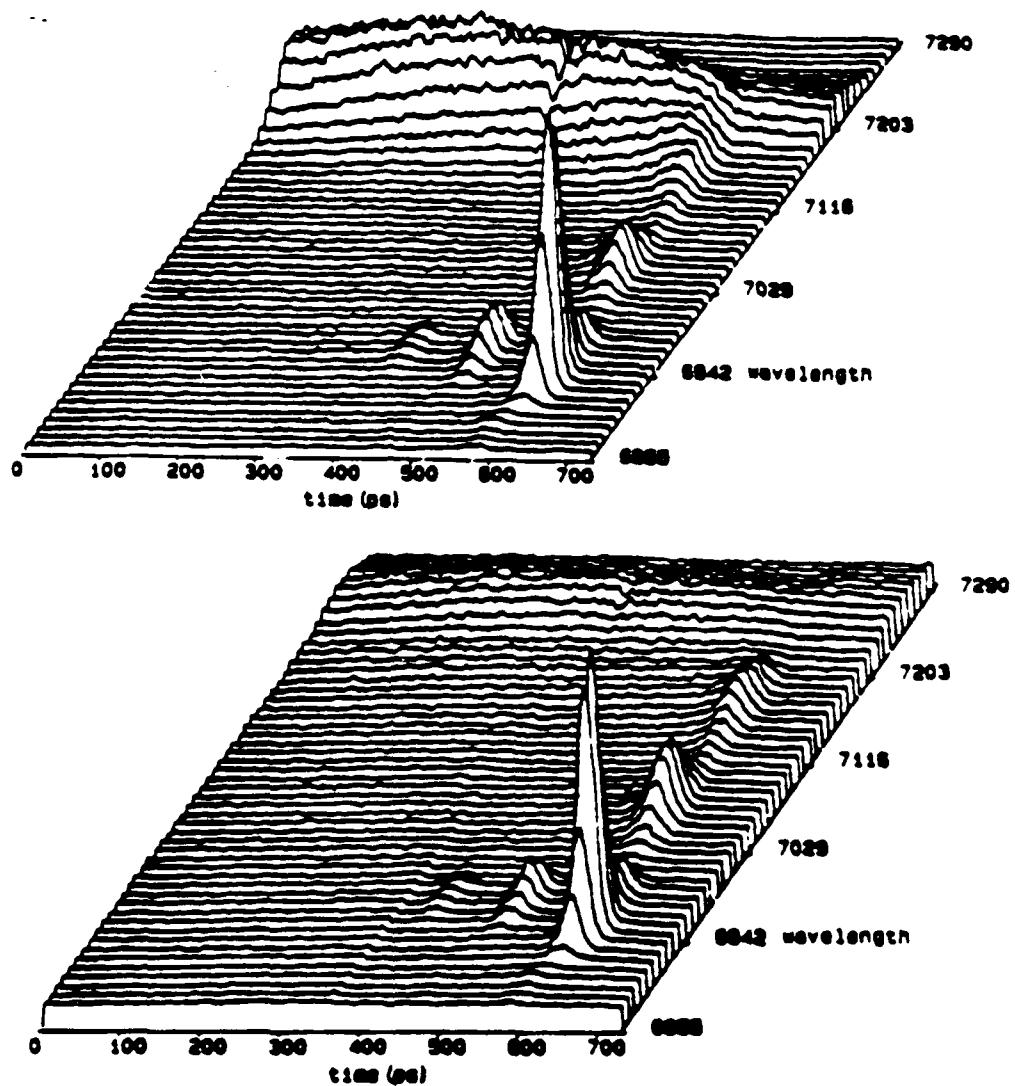


Figure 64: Time-resolved PL spectra of sample A at (a) -3V and (b) -5V bias plotted to show the time-dependence of the different spectral components.

bias, where the  $\omega_2$  and  $\omega_{CT}$  lines are well separated and the  $\omega_2$  rate can be well-defined, the observed  $\omega_2$  decay rate is limited by the streak camera response.

From Fig. 50, the simplest interpretation of the rates is that the  $\omega_1$  line decays with the hole tunneling rate from QW1 to QW2, and the  $\omega_2$  line decays with the electron tunneling rate from QW2 to QW1 (in addition, of course, to the recombination contribution to the total decay rates). The  $\omega_2$  decay is always very fast, because the barrier is thin and the electron mass is low, and at

low bias the electron wavefunction extends over both wells. The  $\omega_1$  decay is at low bias longer than the combination decay rate, but rapidly increases with electric field, until by -3.5V bias the observed decay is limited by the streak camera response time.

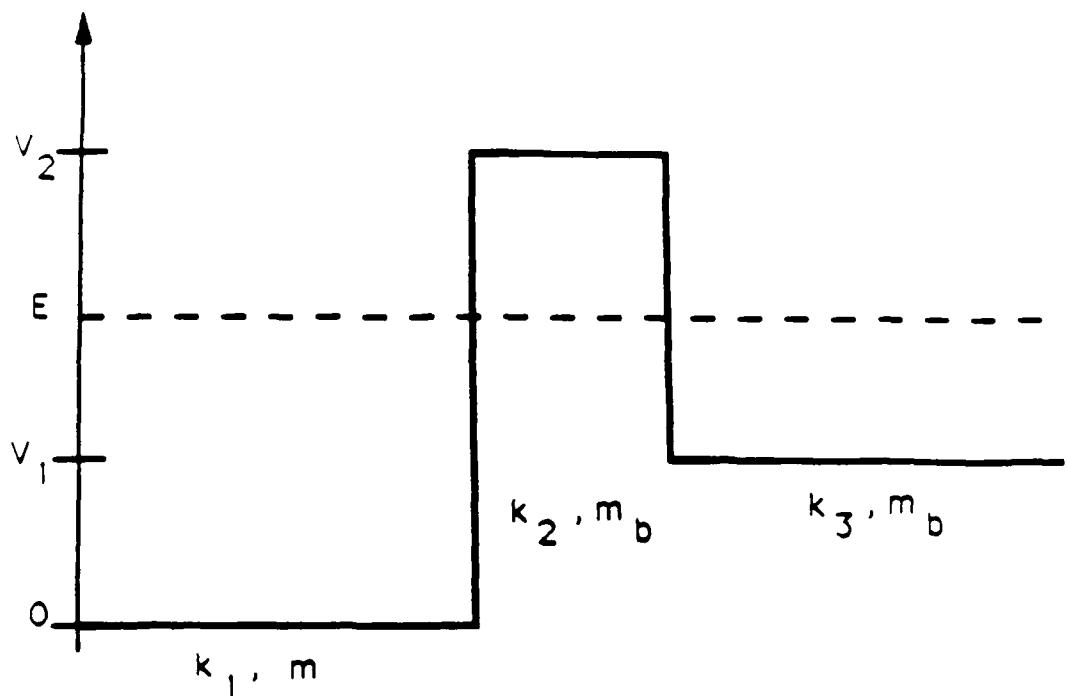


Figure 65: Band diagram for the tunneling-out theory applied to the DQW problem.

An approximation to the rate for a carrier to tunnel from one QW to another may be obtained by applying the theory developed earlier for the tunneling-out problem. A slight modification of the calculation of the transmission coefficients at zero field is necessary due to the misalignment of the band edges of the two QW's; Fig. 65 shows the relevant band diagram. The transmission coefficient at zero field for tunneling from region I to region III is given by

$$T = 4 \left\{ \left( 1 + \frac{m}{m_3} \frac{k_3}{k_1} \right)^2 \cosh^2 k_2 b + \left( \frac{m}{m_b} \frac{k_2}{k_1} \cdot \frac{m_b}{m_3} \frac{k_3}{k_2} \right) \sinh^2 k_2 b \right\}^{-1}$$

where  $m$  is the effective mass in region I,  $m_b$  the effective mass in the barrier,  $m_3$  the effective mass in region III, and the  $k_i$  are the wavevectors in the three regions ( $k_2$  being the evanescent wavevector in the barrier). For the heavy holes tunneling from QW1 to QW2 the  $k$ 's are

$$k_1 = \left( 2m(E - V_1) / \hbar^2 \right)^{1/2}$$

$$k_2 = \left( 2m_b(V_2 - E) / \hbar^2 \right)^{1/2}$$

$$k_3 = \left( 2m_3 E / \hbar^2 \right)^{1/2}.$$

For the electrons tunneling from QW2 to QW1 the  $k$ 's are

$$k_1 = \left( 2mE / \hbar^2 \right)^{1/2}$$

$$k_2 = \left( 2m_b(V_2 - E) / \hbar^2 \right)^{1/2}$$

$$k_3 = \left( 2m_3(E - V_1) / \hbar^2 \right)^{1/2}.$$

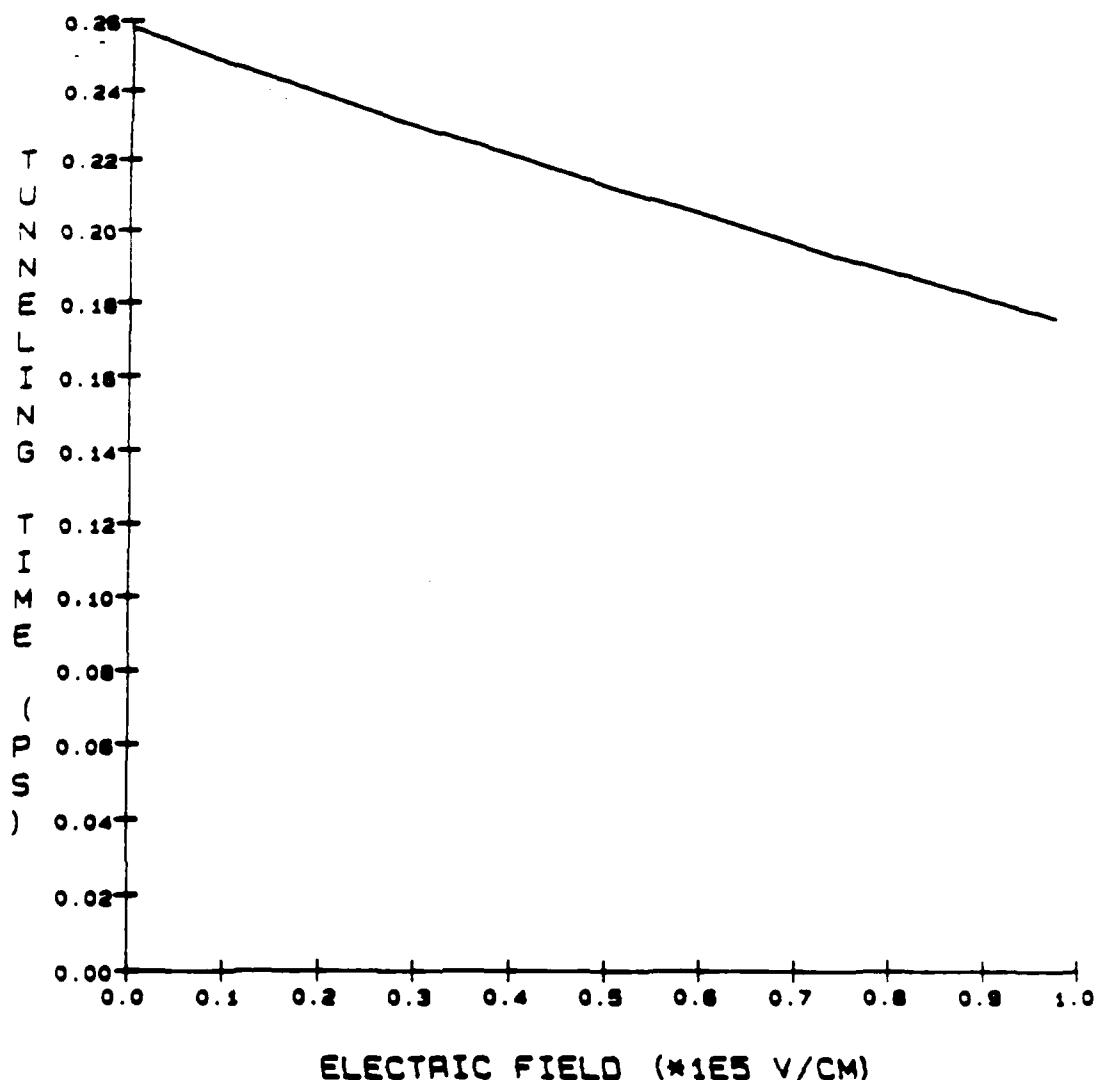


Figure 66: Calculated electron  $QW_2 \rightarrow QW_1$  tunneling times for sample A.

The results for the tunneling rates vs. electric field for a 43-Å barrier, assuming a conduction to valence band offset ratio of 62:38, are shown in Figs. 66 and 67. The electron tunneling rate is much faster than the streak camera response, and so will not be resolved, which is consistent with the experimental results (remembering that at low bias  $\omega_2$  and  $\omega_{CT}$  are indistinguishable). The hole tunneling rate, however, is always much slower than the recombination rate (which is a few hundred picoseconds); thus an interpretation of the  $\omega_1$  decay as the hole tunneling rate requires an anomalously fast decay.

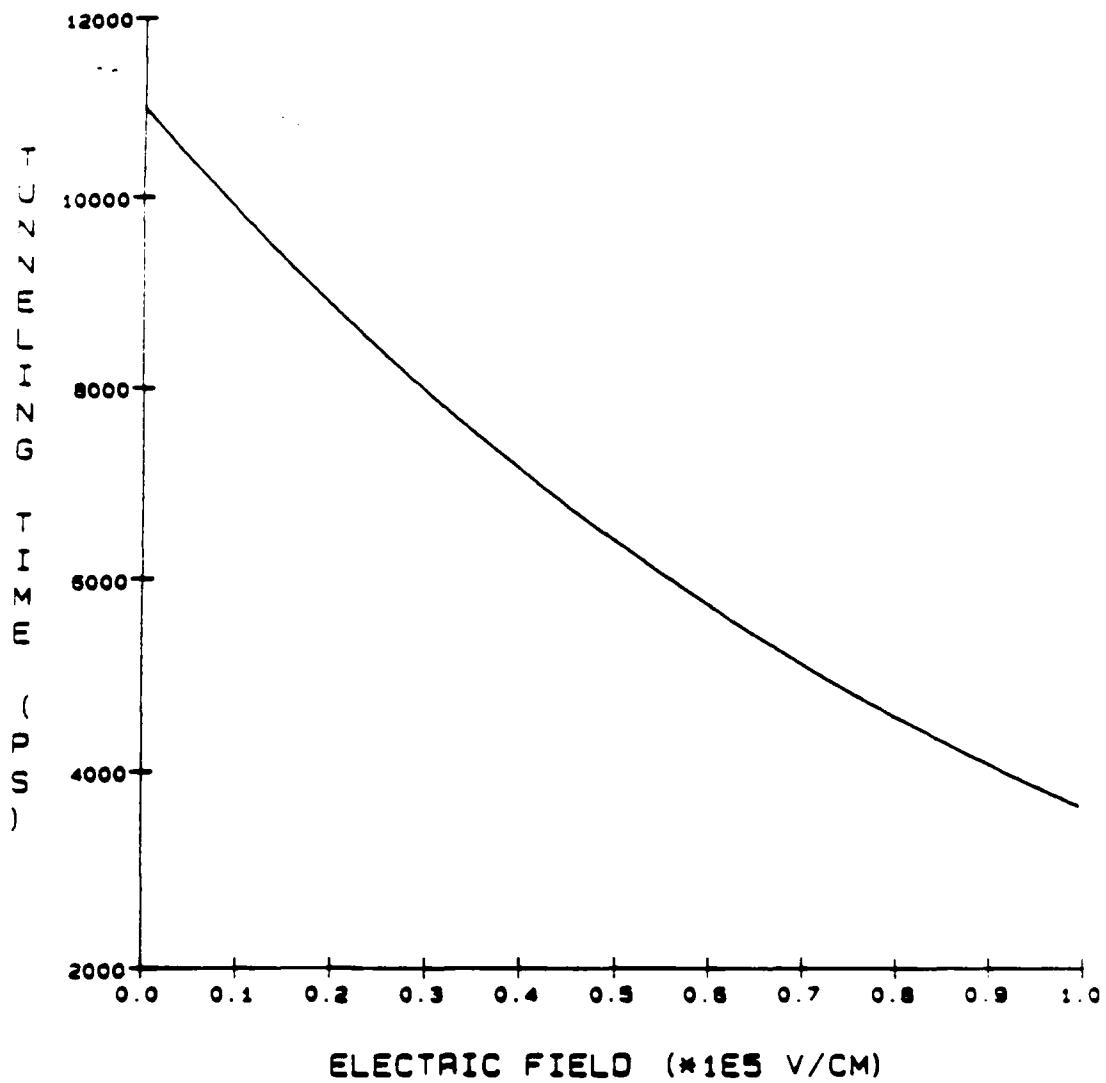


Figure 67: Calculated heavy hole QW1 → QW2 tunneling times for sample A using the tunneling-out theory.

The results for the electron tunneling for the 86-Å barrier are shown in Fig. 68. These calculated rates appear to be too fast to be consistent with the experimental observations. If the tunneling time at high bias were as short as 50 ps, then a substantial charge separation would occur (due to the fact that the holes injected by the laser into QW2 remain there, but most of the electrons would tunnel to QW1 before recombination in QW2 could occur). Hence, the  $\omega_2$  PL intensity would decrease substantially, and a large Stark-shifted component at  $\omega_{CT}$  would appear. This is contrary to experimental observations (see Fig. 56).

From the above argument, it is tempting to conclude that the electron tunneling-out theory does not apply. Before one jumps to this conclusion, however, it is worth repeating the caution that the calculated tunneling rate is an extremely steep function of the assumed barrier height and width. For example, if the barrier is 100-Å thick, the electron tunneling time would be 1 ns at zero field, and would become equal to the PL decay rate only at a field of  $6 \times 10^4$  V/cm. This field was not attained experimentally (due to the screening of the field by the separated charges). Thus the extreme sensitivity of the theory to the growth parameters (and assumed band offset ratio) makes it difficult to conclude to what extent the tunneling-out theory may apply.

The results for the hole tunneling for the 86-Å barrier are not displayed. In this case, the hole tunneling time is never faster than a few hundred microseconds even at high bias; hence, the hole tunneling may be ignored.

Returning to the problem of interpreting the time-resolved PL spectra of sample A, it should be noted that the electrons are injected into the QW's with a small excess kinetic energy. They will, therefore, be warmer than the lattice. It follows that as long as the electron levels in the two QW's are within a few  $kT_e$  of each other, where  $T_e$  is the electron temperature, then the observed decay of the  $\omega_2$  line (in fact, the evolution of the  $\omega_2 + \omega_{CT}$  combined lineshape) will depend on both the tunneling rate and the electron cooling rate in the coupled QW system. We have calculated the initial electron temperatures in each QW. The results are  $kT_e = 5.7$  meV for QW1 and 8.2 meV for QW2 (consistent with the experimental results of Polland *et al.*<sup>7</sup>). Polland *et al.* have shown that these temperatures decay with a time constant of several hundred picosecond in a SQW. When one also includes the effect of the band-filling due to charge buildup in the QW's, it becomes clear the electrons in QW1 can be thermally excited into QW2 over a bias range of several volts, consistent with experimental observation. The fact that the energy resolution of the time-resolved PL system is about 3 meV further complicates the interpretation of the time-dependent  $\omega_2 + \omega_{CT}$  combined lineshape.

A further difficulty with the above interpretation of the  $\omega_1$  and  $\omega_2$  PL decay rates as hole and electron tunneling, respectively, is that it cannot account for the low intensity of the  $\omega_1$  line

unless the hole tunneling has a component that is faster than the streak camera response time of 20 ps. If there were no such fast component, then in fact, the  $\omega_1$  line would dominate the time-resolved spectra, since the population of QW2 would be depleted by electron tunneling, but all the holes would remain in QW1 (at low bias); hence, the instantaneous intensity of  $\omega_1$  would be much greater than the  $\omega_{CT}$  line, which is not what we observe. It is straightforward to show by a simple rate equation analysis that a hole tunneling time of about 36 ps is required to explain the relative intensities of the PL lines in the cw and time-resolved data (with the assumption that the hole tunneling rate is independent of the hole momentum in the plane of the QW). This tunneling rate is several orders of magnitude greater than the simple theoretical estimate given by the tunneling-out theory.

Of course, another interpretation of the relative intensities is that since the QW1 recombination is in  $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$ , which is lower quality material than GaAs, there may be a nonradiative contribution to the decay which dominates. Possibly there are many hole traps, so there are few holes in QW1 with which the electrons may recombine. Another possibility is that the nonradiative recombination rate is so fast in QW1 that the electrons and holes recombine faster than the transfer time of electrons from QW2. The principal difficulty with this interpretation is that for sample B, which was grown in the MBE run immediately successive to sample A under identical conditions, the  $\omega_1$  and  $\omega_2$  lines have roughly equal intensities. It is unfortunately not possible to determine whether the low  $\omega_1$  intensity for sample A is due to fast hole tunneling or nonradiative decay on the basis of PL experiments; the only unambiguous way to settle the issue would be to have a third sample with barrier width between that for sample A and for sample B, so that a trend in relative PL intensities could be clearly observed.

What is clear, however, is that at high fields the decay of both  $\omega_1$  and  $\omega_2$  PL components is streak-camera limited and; therefore, we can conclude that at least at high fields the tunneling is fast (less than 20 ps) for both electrons and holes. The tunneling-out theory cannot account for the fast observed hole tunneling rate.

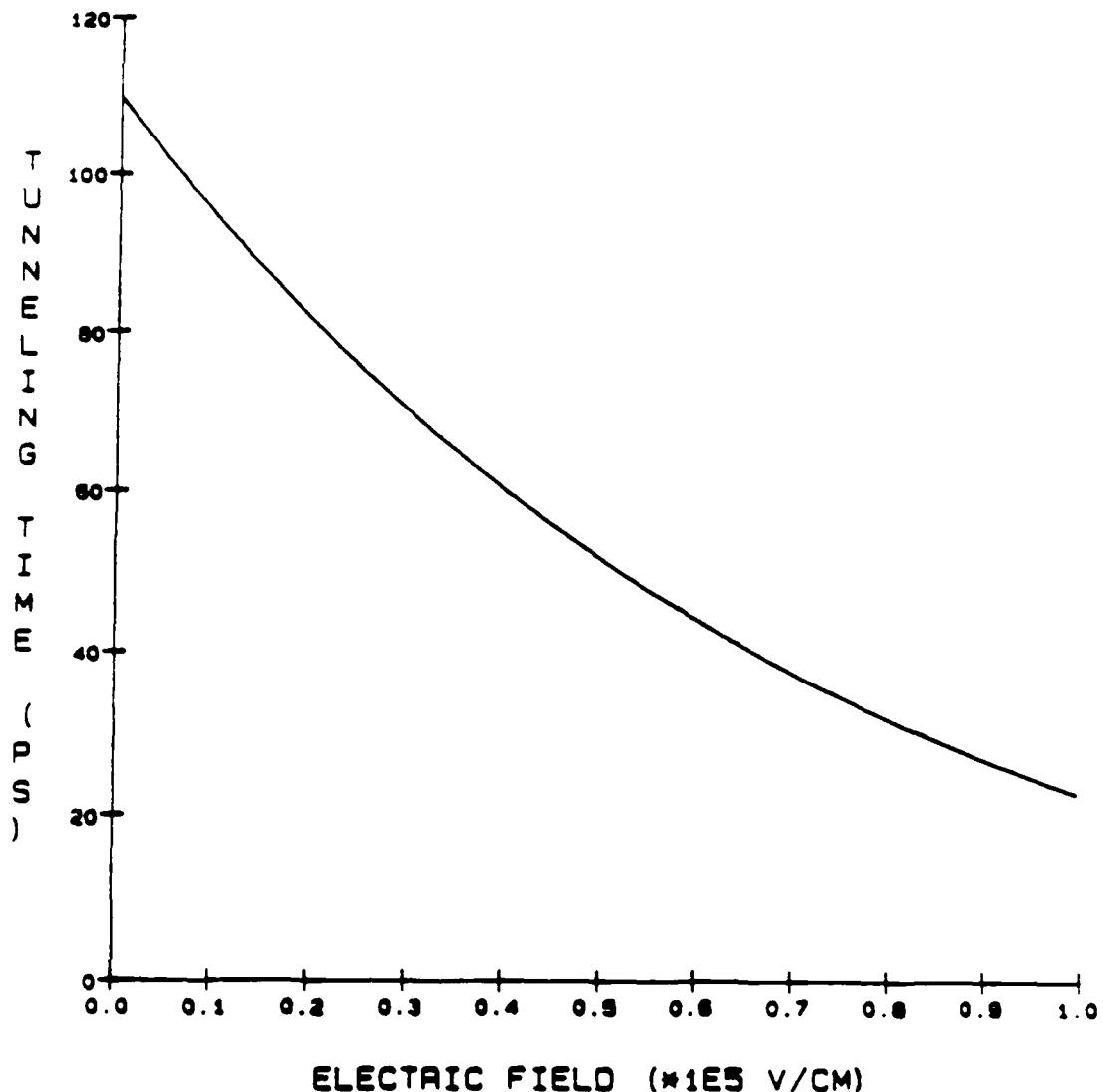


Figure 68: Calculated electron QW2 → QW1 tunneling times for sample B using the tunneling-out theory.

#### IV. RESONANT-TUNNELING DIODE

Resonant-tunneling devices, employing heterostructure quantum wells or superlattices of quantum wells, have been proposed as the next generation of electronics, since they would have gain over very broad bandwidths and be able to attain extremely short switching speeds on account of their very fast transport mechanisms. It is thus appropriate to end this work on ultrafast electrical signals with a chapter that experimentally investigates transport in these unique devices and reports on the fastest electrical switching event yet measured for an electronic component.

The initial development of resonant-tunneling devices to exploit their potentially ultrafast response has taken place in two-terminal devices, or resonant-tunneling diodes (RTD), also known as double-barrier heterostructure diodes (DBD). A heterostructure is a series of junctions formed between two dissimilar semiconductor materials. In the case of the DBD, the device consists of two heterojunctions formed by the presence of one type of semiconductor material grown on both sides of another semiconductor layer with a lower band gap. Careful measurements on this type of device not only provide technological information on its switching operation and potential for high-speed applications, but can also contribute scientific knowledge concerning the mechanisms of transport that occur on the picosecond time scale.

##### IV.A. Theoretical considerations

Resonant tunneling occurs in the double-barrier quantum well when carriers from the Fermi sea of electrons in the bulk semiconductor material on either side of the structure are able to tunnel through one barrier, find an allowed state in the well material, and then tunnel through the second barrier to the bulk material on the other side. The situation of tunneling through a barrier is treated quantum mechanically as a beam of particles impinging on a rectangular region having a potential energy greater than the energy of the electrons. An expression for the transparency of this region, where classically an incident wave would be totally reflected, is given by

$$\frac{1}{T} = 1 + \frac{V^2}{4 E (V - E_e)} \sinh^2 (2ka) , \quad (15)$$

where  $V$  is the barrier height,  $E_e$  is the energy of an electron, and  $a$  is half the thickness of the barrier. The energy difference  $E_e - V$  is the kinetic energy of the electron in the barrier, and its momentum is given by  $hk/2\pi$ . The transmission coefficient is thus proportional to the inverse of  $\sinh^2 ka$ , but even though as the barrier decreases in size, the transmission through it increases,  $T$  for each of the barriers is still only on the order of  $10^{-2}$ . The transmission coefficient for the complete double barrier structure, approximately given by the product of the transmission coefficients for the left and right barriers, is thus extremely small.

However, in a semiconductor there are no bulk states in which electrons may exist, but rather there are allowed energy bands with forbidden gaps between them. A single layer of semiconducting material grown between two layers of another semiconducting material, where the

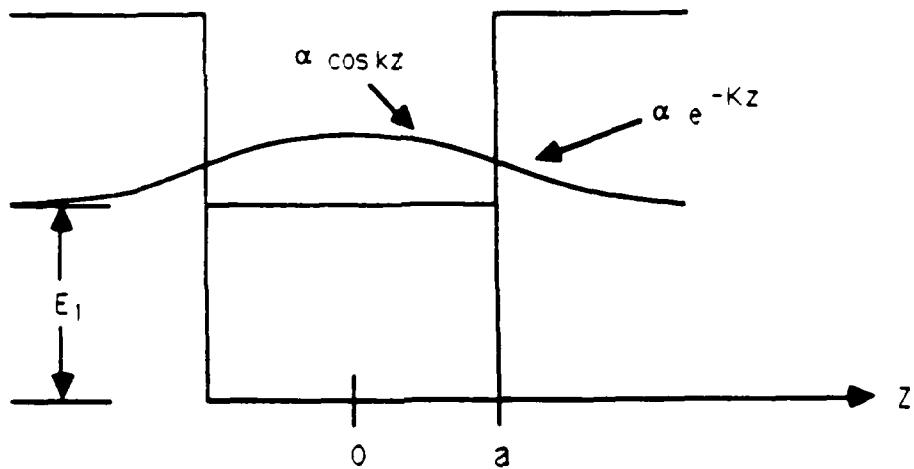


Figure 69: Bound energy levels and electron wavefunction in a finite double-barrier quantum well.

embedded layer has a lower bandgap than the surrounding layers, would also exhibit these allowed, or bound, energy states. In the case of an infinitely deep well of finite extent, the wave function of an electron in one of these states is found to be zero in the confining layer.<sup>90</sup>

However, for a finite quantum well, an electron's wave function is not absolutely confined between the barriers, but rather it is discovered to decrease exponentially in the barrier regions due to the evanescent states in each of the layers.<sup>91</sup> This situation is depicted in Fig. 69, where  $k$  and  $K$  are momenta terms for the electrons in the two layers. It should be noticed that as the size of the structure decreases, it is possible for the electron wave function to be transmitted through the entire double barrier structure. Since this can only occur when an electron in an emitter on the other side of a barrier from the well is at the energy of a state in the well, the condition is considered to be one of resonance, and hence the term resonant tunneling.

The global transmission coefficient for the finite double barrier quantum well structure has been derived and is reproduced<sup>92</sup> below in order that the resonant aspect of this phenomenon may be further clarified:

$$T_G = \frac{C_0}{C_1 T_l T_r + C_2 \frac{T_l}{T_r} + C_3 \frac{T_r}{T_l} + C_4 \frac{1}{T_l T_r}}, \quad (16)$$

where the subscripts  $l$  and  $r$  identify the left and right barriers, and the  $C_x$  coefficients are phase factors that to first order may be considered to be constant and of the same order of magnitude. As discussed previously, the magnitudes of the individual barrier transmission coefficients are considered to be much less than unity, so when the energy of an incident electron does not correspond to an energy state in the well, the global transmission coefficient reduces to

$$T_G = \frac{C_0}{C_4} T_l T_r = T_l T_r. \quad (17)$$

The global transmission is small, and the well would have an effect only on the phase of an electron. If the phase factor  $C_4$  goes to zero, however, as happens to be the case when an electron

is at an energy corresponding to a level in the well, the global transmission coefficient peaks:

$$T_G = \frac{C_0}{C_2} \frac{T_r}{T_l} \quad \text{or} \quad T_G = \frac{C_0}{C_3} \frac{T_l}{T_r} . \quad (18)$$

where the  $T$  in the numerator would represent the barrier which exhibited the smaller transmission coefficient. This corresponds to the resonant condition described earlier, and depending on which barrier had the maximum and minimum transmission coefficient, the global expression could be written as

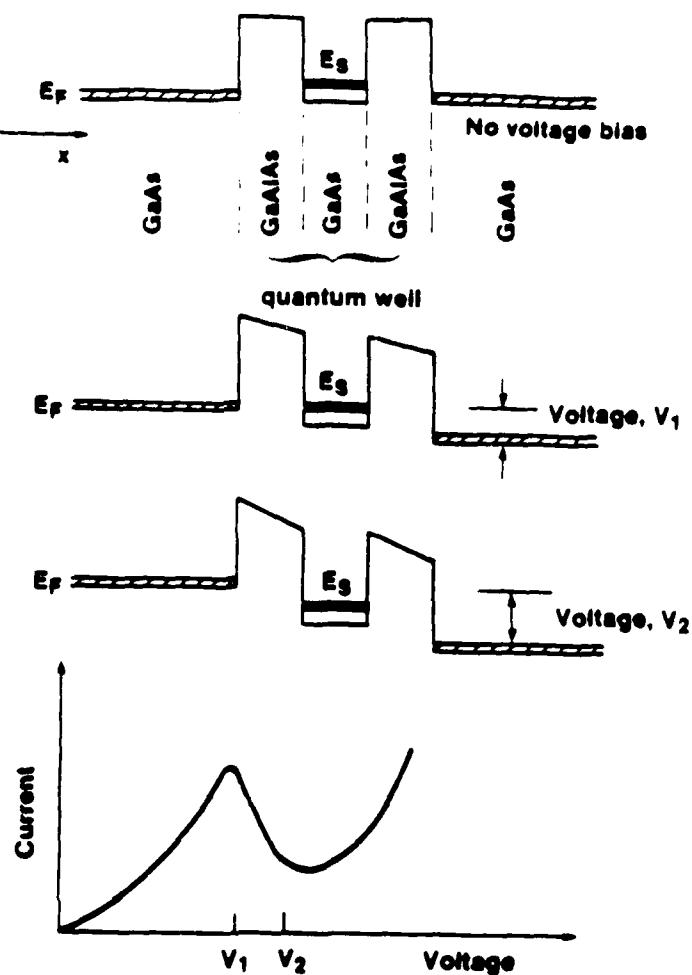
$$T_{G, \text{res}} = \frac{T_{\text{min}}}{T_{\text{max}}} . \quad (19)$$

This demonstrates that even if the individual barrier transmission coefficients are very small, as long as the energy of an electron matches that of the bound state, it has an excellent chance to be transmitted through the entire structure.

#### IV.A.1. Current-Voltage Characteristic

The current-voltage (I-V) characteristic of a tunnel diode is interesting and useful due to its negative differential resistance (NDR) region. The RTD also exhibits an NDR region, and thus may eventually find applications as an oscillator, mixer, and fast switch. An explanation for the existence of NDR in this device follows, referring to Fig. 70. The layers that the double barrier structure are embedded

between are heavily doped semiconductor (GaAs doped to  $10^{18} \text{ cm}^{-3}$  typically) material, and they act as the contact electrodes to the RTD, although there would actually be ohmic metal contacts outside of these heavily doped regions, too. A voltage bias is applied across the RTD through the use of these contacts. The equilibrium, zero-bias case is shown at the top of Fig. 70. Only the



210

Figure 70: Schematic band diagram and current-voltage characteristic for a hypothetical resonant-tunneling diode.

first allowed state in the well material is given, and the Fermi level is indicated by the shaded area in the bulk material. At the very bottom, the I-V curve indicates the amount of conduction for the given bias voltage, where at  $V = 0$  volts the current is also at the origin.

As the bias voltage is applied to the RTD, the resonant state is pulled down in energy with respect to the more negative electrode, and when the Fermi level in this electrode is at the same energy as the bound state, enhanced conduction occurs and the I-V characteristic peaks. In our schematic drawing, this has happened at  $V = V_1$ , which can also be given by  $V = 2 E_1/e$ , since the

applied voltage is split into two equal voltage drops at each barrier. When the resonant state is pulled down below the conduction-band edge of the heavily doped material, the tunneling current decreases. The resonant state is well below this point at a voltage of  $V = V_2$  in the figure, where the current is seen to have attained a minimum. At bias voltages greater than this value, there is significant current flow of electrons from the Fermi sea that have energies above the top of the potential barrier, and the characteristic in this region appears similar to that for a p-n junction as it begins to conduct. The rise in conduction to the peak of the I-V curve is reasonably gradual at room temperature, as the tunneling current depends on the density of occupied states in the electrode as determined by the Fermi function. This peak should therefore become quite sharp at very low temperatures. Of course, if there are more allowed states in the well, the I-V characteristic would have multiple peaks, and these have been observed by a number of investigators.<sup>95</sup>

#### IV.A.2. Coherent vs Sequential Resonant-Tunneling

The global transmission function at the resonance condition, as described earlier in the chapter, is observed to achieve a maximum when the transmission coefficients of the two barriers are equal to each other. This transmission exhibits a resonant enhancement that behaves much like that in an optical Fabry-Perot resonator. This assumes that there is negligible scattering of the electrons in the well, which is not actually the case in the vicinity of room temperature, but even at non-cryogenic temperatures there is some enhancement in transmission due to the resonance.

With the barriers behaving like the mirrors of a cavity resonator, the electrons incident on the well from the emitter at the energy of the bound state experience multiple reflections in the gap between the barriers. When a small amount of the wave function leaks out of a barrier upon each reflection, it is in phase with the other transmitted waves (they exhibit phase coherence), and out of phase with the reflected waves. From the theory of the resonance properties of passive optical cavities, it is known that the amplitude of the electron wave function circulating between the barriers would build up inside the well, and, depending on the transmission coefficients of each of

the barriers, either enhance the transmitted waves or the reflected ones. As indicated in Sect. IV.A, when the transmission coefficients of the barriers are matched, it is the transmitted waves that are enhanced and the reflected ones that are cancelled on resonance. This is analogous to an impedance matched condition on an ordinary transmission line where the load and characteristic impedances are the same so that there are no reflections.

The initial belief was that the peak in the transmission of current through the double barrier structure was entirely due to this coherent tunneling process that occurs when electrons in the Fermi level are aligned in energy with the bound state. Theoretical calculations have consistently proven to overestimate the peak-to-valley ratios of the I-V curve, however,<sup>53</sup> with a logical conclusion being that as an electron scatters off other particles in the well material, the collisions randomize the phases of the electron waves and prevent the build up of the amplitude of the wave function in the well due to the multiple reflections. Since there would be no resonant enhancement of the transmission, the electrons would tunnel through the region without maintaining the coherence of the incident wave, sequentially scattering until they reached the extent of the structure at an amplitude lower than that expected from the resonance condition.<sup>94</sup> While this would indicate that coherent resonant tunneling was present just because there was a peak in the I-V characteristic, it would also substantiate the 93 of incoherent resonant tunneling due to the diminished peak in the conduction.

Further theoretical work, however, has suggested that although the scattering destroys the phase coherence and strongly reduces the probability for coherent tunneling, the tunneling current does not depend on whether the resonant tunneling is coherent or sequential, but only on the alignment of occupied states in the emitter with the bound state. This is an indication that it may be necessary to concentrate on the time response of the tunneling process rather than the dc studies of the I-V curve in order to determine which tunneling mechanisms dominate, or whether there is actually any contribution from coherent resonant tunneling to the current at all.

A study of the time it takes an electron to traverse the double barrier structure is useful because there is a difference between the build-up time of the amplitude of the wave function in the

well due to the multiple reflections of the electron and the time for the electron to scatter across the region. The duration of time before the interference produces a high transmission on resonance is on the order of  $h/(2\pi\Gamma_p)$ , where  $\Gamma_p$  is the FWHM of the energy interval over which the resonance occurs. This time,  $\tau_p$ , can be considered to be a quantum mechanical time delay for the onset of the coherent resonant tunneling current. The energy line widths for several RTDs with different AlGaAs-barrier thicknesses and a 5-nm GaAs well width are given in Table IV.1 at the end of the chapter,<sup>93</sup> along with the approximate time constant for the build-up time for the coherent wavefunction at the output of the device. These times are seen to vary from over 50 ps for the barrier width of 70 Å to about 100 fs for the barrier width of 20 Å.

When collisions in the structure interrupt the development of the coherence of the wavefunctions, the resonance does not occur. In terms of timing, the scattering time, or time between collisions,  $\tau_s$ , is less than the build-up time of the cavity, and thus the probability is greater that a collision occurs than that interference between wavefunctions occurs. These collisions may be elastic, such as scattering between carriers, carriers and impurities, or carriers and poor layer interfaces, or they may also be inelastic, such as when phonons carry away energy towards the other side of the structure. These collisions provide alternate channels for current that may compete with the coherent transport. Regardless, the total resonance energy width of the system,  $\Gamma_T$ , becomes equal to  $\Gamma_s + \Gamma_p$ , so that the resonance is said to be collisionally broadened. The total device response time,  $\tau$ , is thus observed to become shorter on account of scattering contributing to the transport of the carriers before the slower, coherent transport has a chance to dominate. Calculations on the example AlGaAs/GaAs RTDs<sup>93</sup>, which have a mobility of about  $7 \times 10^3 \text{ cm}^2/\text{s}$ , indicate that  $\Gamma_s$  broadens the resonance energy width of any of the structures in Table V.1 by approximately 2 meV at room temperature. This does not have a great effect on the RTD with the 20 Å barriers, as it decreases the response time from about 100 to 75 fs, but it makes a large improvement in the response time of the RTD with the 70 Å barriers so that its tunneling

time decreases by over two orders of magnitude:

$$\tau \equiv \frac{\Gamma}{\Gamma_r + \Gamma_s} = \frac{6.58 \times 10^{-16} \text{ eV} \cdot \text{s}}{0.0128 \text{ meV} + 2.13 \text{ meV}} \approx 300 \text{ fs} .$$

Therefore, at room temperature, the response always appears to be subpicosecond for the example RTD, either because of the small barriers and large coherent tunneling linewidth, or because of the collisionally broadened resonance.

At lower temperatures, such as 77 K for a liquid nitrogen bath, scattering can be frozen out and markedly decreased so that the mobilities may exceed  $10^5 \text{ cm}^2/\text{s}$  and the time between collisions rises above 1 ps. The broadening now would be less than 0.5 meV, which would decrease the response for the 20-Å-barrier RTD by only by about 8%, but would still diminish the tunneling time for the 70-Å-barrier device to between 1 and 2 ps. An observation of a response on the order of a single picosecond from the former device would signal the definitive presence of coherent resonant tunneling, while the picosecond response in the latter RTD, rather than the 50 ps expected if there were no broadening, would lead to the conclusion that sequential tunneling still dominated. The connection between this tunneling time and the actual observed operation of an RTD are discussed in the next section.

### V.B. Tunnel Diode Operation

Part of the usefulness of the I-V characteristic of any device is that the operating behavior of the component may be deduced from the intersection of this curve with a load line. The load line is simply a plot of the current through the load versus the voltage across the device, and so for resistive loading it is a straight line whose intercept with the voltage axis is the input voltage to the circuit and whose intersection with the I-V curve gives the operating voltage and current. It is particularly useful when applied to drastically non-linear devices such as the tunnel diode. By placing load lines on the I-V curve, the various useful operating regimes of the tunnel diode can be

identified in Fig. 72. A potential dc load line may be considered whereby the absolute value of the negative resistance is greater than the series resistance present in the circuit. If the circuit containing the diode can be tuned so that the parasitic circuit parameters attain such values that the impedance  $L/RC$  is equal to the magnitude of the negative resistance, then stable sinusoidal oscillations are output from the circuit. Qualitatively, it should be remembered that all conventional circuit components have positive resistance and thus dissipate power; but when a tunnel diode contributes its negative differential resistance to a circuit so that the net resistance vanishes, no power loss results and the circuit oscillates at its resonant frequency. If the magnitude of the tunnel diode's negative resistance is greater than the impedance, then the load line is steeper than the slope of the NDR region and the oscillations are damped. However, when this magnitude is less than the impedance, such as when the circuit has a large inductance and a small capacitance, the slope is more shallow than the NDR's slope and a bistable, or switching, mode is realized.

Bistable operation results when there are multiple intersections of the load line and the I-V curve. In Fig. 71 there exist three intersections, two of which are stable, and one, point 3, which is considered to be in a region of unstable equilibrium. If a circuit is brought to this point, the current immediately begins to change until it settles at either point 1 or point 2, the two stable operating states of the circuit. The circuit may be switched from one of the points to the other by the application of a voltage. This process serves to change the intercept of the load line with the voltage axis, but since the impedance does not change when this occurs, the load line is simply shifted to a position parallel to where it started. For instance, if the circuit begins at point 1, and a voltage is applied such that the load line shifts to exceed its intersection with the current peak, the operating point shifts to the other branch with positive slope until the voltage is removed and the load line shifts down until it intersects the I-V curve at point 2. The speed with which this switching action takes place and the relevance that this has to the tunneling mechanisms are to be discussed presently.

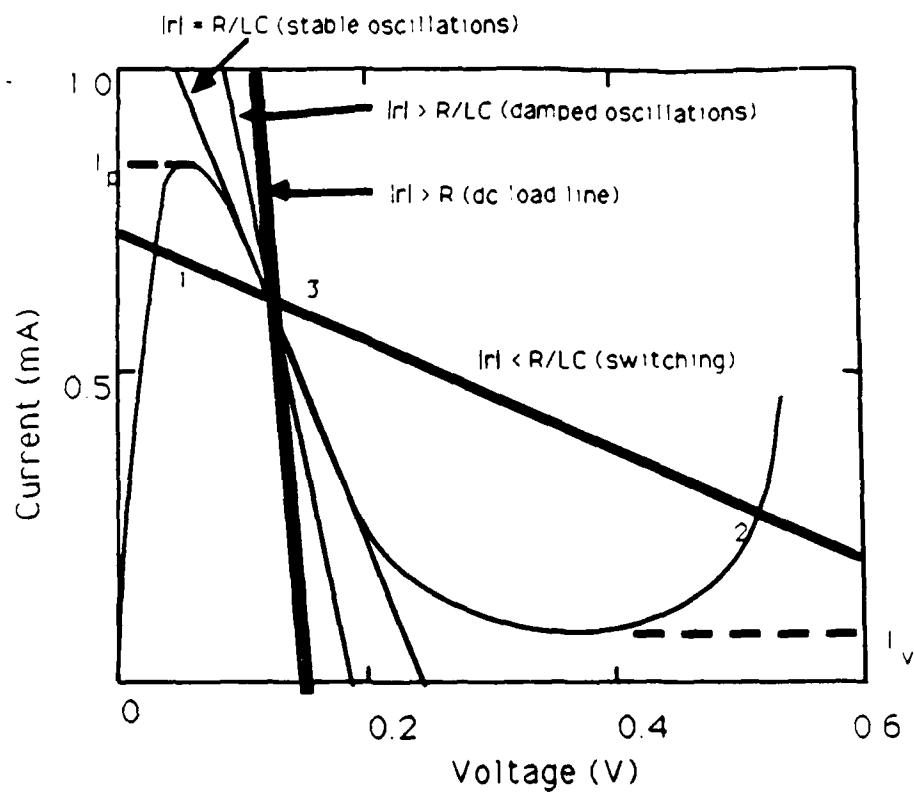


Figure 71: Hypothetical resonant tunneling diode I-V curve with load lines indicating operation as switch and oscillator.

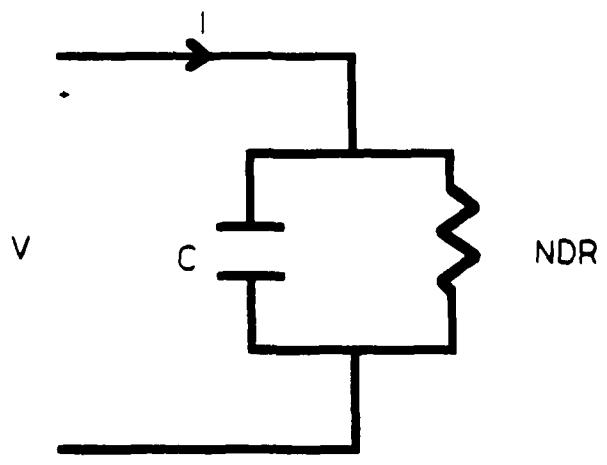


Figure 72: Simplified resonant tunneling diode equivalent circuit.

The switching speed of the RTD is constrained due to the fact that for any logic application, or even any measurement in a practical situation, the device exists in a circuit. A simplified version of the RTD equivalent circuit, a device capacitance,  $C$ , in parallel with a resistance that includes the negative differential resistance, is displayed in Fig. 72. The switching would occur when a current step,  $\Delta I$ , is furnished to the RTD so that the load line exceeds the peak of the I-V curve. When the capacitance is considered to be constant during the switching, the switching speed from points 1 to 2 in Fig. 71 only depends on the time it takes to charge the capacitor from the voltage corresponding to the point marked 1 to the voltage at the point marked 2.<sup>94</sup> The rise time of the switching event can be estimated as

$$t_r \equiv C \frac{(V_2 - V_1)}{(I_p - I_v)}, \quad (20)$$

where  $V_2$  and  $V_1$  are the voltages corresponding to the points 1 and 2, and  $I_p$  and  $I_v$  represent the peak and valley currents as indicated on the I-V curve in Fig. 71.

More accurately, a load resistance,  $R_l$ , would also be included in the circuit, so that the  $\Delta I$  would have to be large enough so that it was greater than the peak current plus the peak voltage divided by the load resistance, or if there is a constant bias with the step superimposed on top:

$$I_B + \Delta I > I_p + \frac{V_p}{R_l}, \quad (21)$$

in order to cause switching in the RTD circuit. It would then be more accurate to write the expression for switching time as<sup>94</sup>

$$t_r = C \int_{V_1}^{V_2} \frac{dV}{I_B + \Delta I - \frac{V}{R_l}} = C R_l \ln \frac{I_B + \Delta I - \frac{V_1}{R_l}}{I_B + \Delta I - \frac{V_2}{R_l}}, \quad (22)$$

where  $V$  is the voltage across the RTD. Typical values for the rise time of the switching event can be on the order of 100 fs if the capacitance is not large and the peak current can be kept in the milliamp range for a large load resistance. More commonly the switching time is in the single-pi-

cosecond regime, but if a large device capacitance is present, switching takes place in the range of tens of picoseconds.

In the experiments conducted on the RTD, the physical quantity measured has been the time for the current to switch from a high value at the peak of the I-V curve to a lower value in the valley. As just discussed, this is not necessarily the same as the tunneling or response time of the RTD, due to the presence of the device capacitance in the circuit. In this sense the capacitive effects are the ones that are measured, so that the switching time is an upper limit of the tunneling time. That is, we measure a time that the tunneling time could not exceed, since if the tunneling time were longer, it would be measured instead of the parasitic response. If, because of the presence of a narrow resonance energy width, the tunneling time were to be on the order of tens of picoseconds, and the capacitance of the device set a lower limit on the measurement response to a time on the order of a single picosecond, then the response measured would surely be the one resulting from the tunneling.

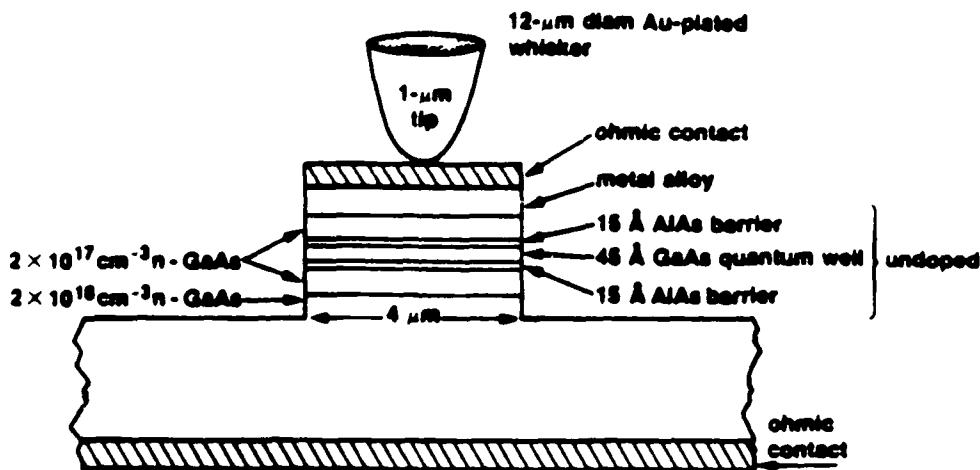


Fig. 73. Cross-section of AlAs/GaAs resonant-tunneling diode.

#### IV.C. Switching-Time Measurement

The resonant tunneling diodes used in the experiment discussed here were developed and fabricated at the Lincoln Laboratory of the Massachusetts Institute of Technology. They were

grown by molecular beam epitaxy to have two barriers of 1.5-nm-thick aluminum arsenide (AlAs) separated by a GaAs well 4.5 nm thick. The outer regions of the GaAs contained silicon doping to an electron density of  $2 \times 10^{17} \text{ cm}^{-3}$ . These material parameters resulted in a peak-to-valley ratio at room temperature for the I-V curve of 3.5/1, and the peak current density was  $4 \times 10^4 \text{ A/cm}^2$ . Although grown on a large substrate of GaAs, the tiny device is isolated on a mesa of 4-mm diameter. This is depicted in Fig. 73, where an ohmic contact of gold/nickel/germanium resides on a layer of  $n^+$ -GaAs on the top of the mesa, and another ohmic contact is present on the back of the heavily doped substrate material. Since the charge storage in the device occurs predominantly in the areas just outside of the double-barrier region, by making the cross-sectional area small, the capacitance can also be made small.

The barriers were fabricated of AlAs in this instance rather than AlGaAs to maximize barrier heights up to  $\sim 1.0 \text{ eV}$ .<sup>95</sup> This reduces the thermionic emission of electrons over the top of the barriers and can therefore increase the peak-to-valley ratio. The AlAs barriers could also be grown at the relatively low temperature of  $560^\circ\text{C}$ , helping to yield smoother heterojunctions and fewer traps.<sup>95</sup> The barriers have been made to be only 1.5 nm thick mainly for reasons associated with the analog applications of the device. The current density is greater for thinner barriers, and the higher current density results in a higher output power for the device when applied as an oscillator. The doping density of the bulk GaAs regions just outside the double-barrier structure has been chosen to help minimize the capacitance while maintaining the GaAs as an emitter. Due to the presence of the potential barriers, charge accumulates on the left side of the structure and is depleted on the right side. A lower doping density reduces the capacitance by increasing the width of the depletion region and lengthening the distance between the plates of the capacitor. The depletion width into the doped collector region was approximately 70 nm, and using a relative permittivity for GaAs of 13.1, the capacitance was found to be about 20 fF, or the same as the measured value. The series resistance was measured to be 15W.

The estimate of the rise time from Eq. IV.8 returns to the more simplified estimate of Eq. V.B.1 when the load resistance is large with respect to the voltages  $V_1$  and  $V_2$ , as is the case in

this experimental work. The values of the voltages before and after switching, as well as the peak and valley currents, have been measured from a curve tracer in order that we may get an idea of the switching time. These measurements gave  $V_2 - V_1 = 0.44$  V and  $I_p - I_v = 3.0$  mA, and the rise time was thus computed to be 2.9 ps.

#### IV.C.1. Test Environment

With switching times on the order of a single picosecond, it once again is necessary to use the electro-optic sampling system in order to resolve an electrical event of short duration. Instead of the signal at the output of the photoconductive switch being guided onto a transmission line, it is coupled into the device and then the output of the device appears on the transmission line. The test geometry is demonstrated in Fig. 74, where the two optical beams and the electronic connections between the electrodes and the device are shown. A GaAs wafer with coplanar electrodes, one with a 20-mm long switch gap near one end, was used as the source of ultrafast-rise-time electrical impulses to the device. The sampling crystal also had a coplanar stripline with 20-mm-wide lines separated by 20 mm deposited on it. The device wafer, diced into small chips of approximately  $0.025 \times 0.025 \times 0.010$ -cm size, contained an array of tunneling diodes on one of the larger faces, and an ohmic contact on the other large face. The chip was mounted on the  $\text{LiTaO}_3$  sampling crystal on one of its sides so that the array of mesas faced back towards the switch, as shown in Fig. 74. The back of the chip then was connected to an electrode on the sampling crystal using highly conductive epoxy - Epo-Tek H20E silver-filled epoxy with  $\rho = 100$  mW-cm. The sampling beam passed between the electrodes of the bulk electro-optic sampling crystal in order to measure the field on the transmission line at the output of the device. A wire bond connected the ground electrode from the GaAs to the  $\text{LiTaO}_3$ , and a provision was also made for a dc bias to the diode through a looping, inductive wire bond that connected to the electrode on the output side of the switch. The electrical output from the sampling crystal was coupled to SMA connectors and through coaxial cable so that it could be monitored on an oscilloscope.

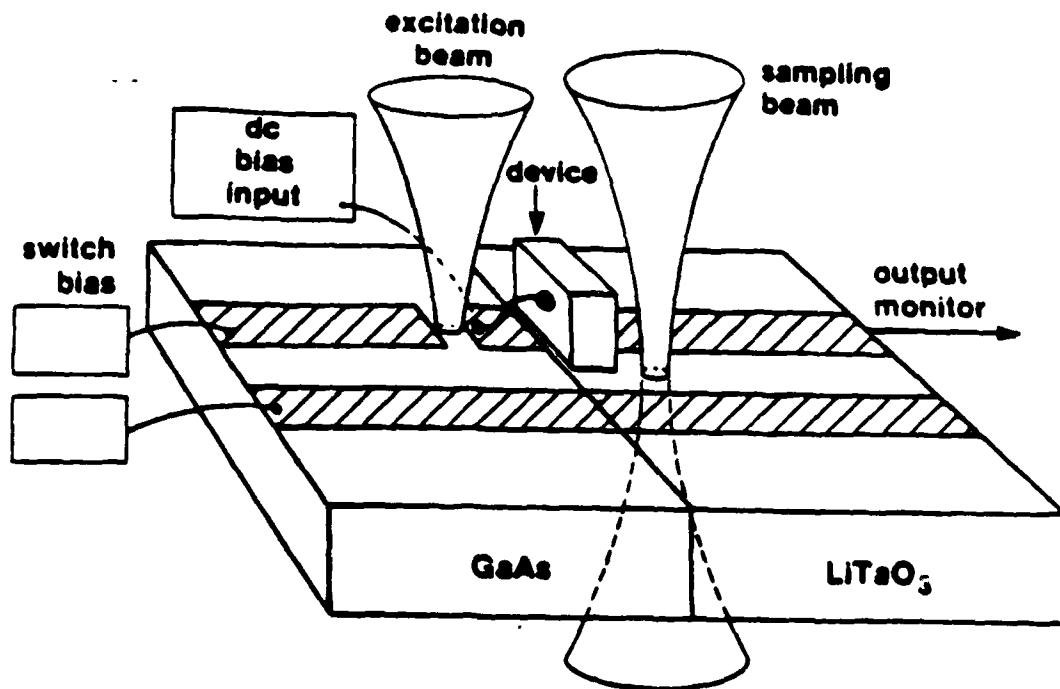


Figure 74 Experimental test geometry for electro-optic sampling of signals from the resonant-tunneling diode.

The contact to the top of the mesa containing the diode was made using the only technique for connecting externally to a pad of such small area - a whisker wire.<sup>96</sup> The whisker was a 12-mm-diameter wire of phosphor-bronze, or C-bronze, which was sharpened down to approximately a 1-mm-diameter at its tip and then plated with gold in order to reduce its resistivity to less than 10 mW-cm. The wire was sharpened using an electro-etching technique where the tip was dipped ~2 mils into an 8% solution of sulfamic acid ( $H_3NO_3S$ ) so that the meniscus of the fluid extended slightly up the wire. Electrodes were connected to the pin vice holding the wire and a copper shim acting as the cathode was inserted into the acid bath about one inch from the wire. At that point a clean, debounced switch was closed for 1-2 seconds so that 18 V was applied across the solution (which was heated to 30° C), and the wire was etched at a rate proportional to the amount of acid surrounding it so that the tip became pointed. The electro-plating of gold onto the whisker was accomplished using a Sel-Rex Pur-A-Gold solution, heated to ~30° C and stirred magnetically.

The entire length of the whisker to be used was now inserted into the fluid, and 2.5 mA of current was supplied through the plating solution with the whisker as the cathode for about 90 seconds.

The whisker then had to be bent in order for it to be able to make contact with both the electrode at the output of the switch and the mesa that resided within 250 mm above the surface of the sampling crystal supporting the diode chip. It was also necessary to have the whisker bent so that it was flexible when it made contact with the diode mesa. By maintaining its flexibility, a strong contact that stayed in place could be made.

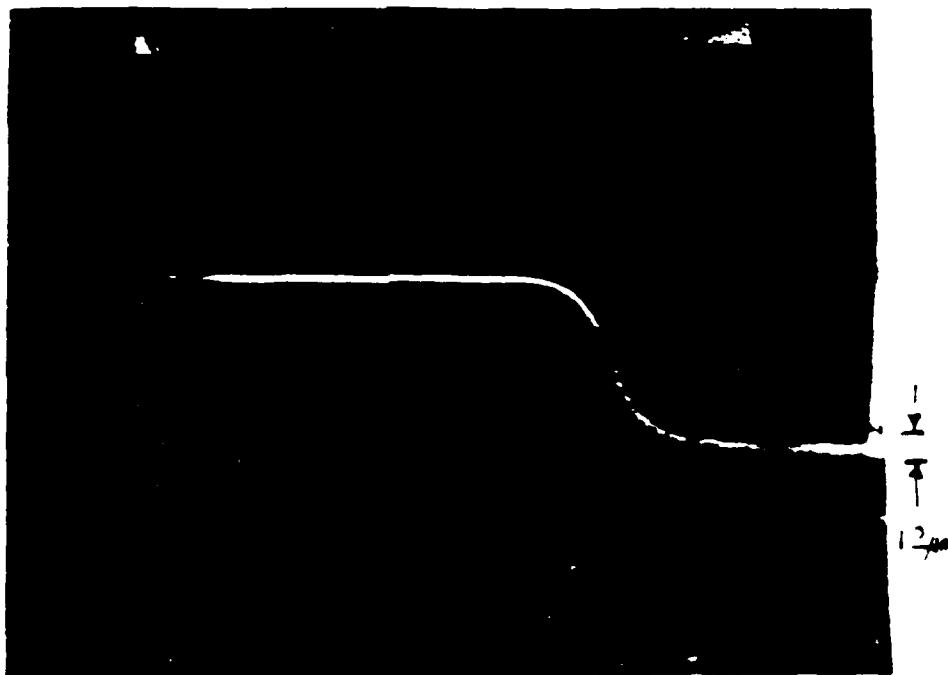


Figure 75. Phosphor-bronze whisker wire for contacting to resonant-tunneling diode. The 12-mm-diameter wire is electro-etched to a 1-mm-diameter at its tip. (Photo courtesy NASA Jet Propulsion Laboratory.)

using the whisker. Once a measurement of the distance from the switch to the diode was made, the whisker was bent between two tungsten electronic probe tips. The procedure was observed at a magnification of 180X under a stereo-microscope. The whisker was held fixed parallel to the

ground, while the probe tips were manipulated using X-Y-Z translation stages. By placing one probe next to the whisker and moving the other probe so that it caused the whisker to bend around the first probe, two bends were made in the whisker in order to attain a rough S-shape that would still be pliant when a force was applied to the tip. A picture of the shaped whisker is shown in Fig. 75. The bent whisker, soldered to a metal support post, was then clamped into a pin vice on another X-Y-Z translation stage, and the whisker and diode array were viewed through a magnification as great as 400X. The whisker tip was translated toward the diodes until it was just tens of microns away, and then the whisker was lowered until it touched the electrode on the GaAs. The final distance from the whisker's tip to the mesa was then covered as the bottom of the whisker was dragged along the electrode, and the tip was placed upon the mesa. The contact at this point was still rather dubious, so the whisker was moved several more microns longitudinally until the force from the spring created in the bends of the wire ensured a good connection. A sharpened whisker with its tip contacting a diode mesa on the face of the RTD chip is seen in Fig. 76.

To contact the other end of the whisker to the circuit, the length of the wire resting on the metal electrode was covered with Chomerics Cho-Solder 574 silver epoxy ( $\rho = 4 \text{ mW-cm}$ ), and left under a heat lamp at  $T = 100^\circ\text{C}$  for 15 minutes. When the epoxy was completely cured, the wire was cut away from its support post by a scalpel blade positioned using an X-Y-Z translator. Occasionally, the whisker wire moved off the mesa at some point after the epoxy had cured, so the tip was repositioned on a diode by manipulating the whisker with the tungsten probe tips again. The use of the 400X magnification was only necessary when it was desired to view the exact mesa the whisker would contact. It was more common to use 180X magnification and move the whisker until the resistance between the dc bias line and the output electrode on the sampling crystal was observed to drop to a low value.

The test geometry from Fig. 74 can also be viewed as an equivalent circuit, as in Fig. 77. The RTD is represented as a capacitor in parallel with a variable resistor,  $G$ , which can be negative, and these are in series with a positive resistor,  $R_s$ . The RTD was connected in series with a load

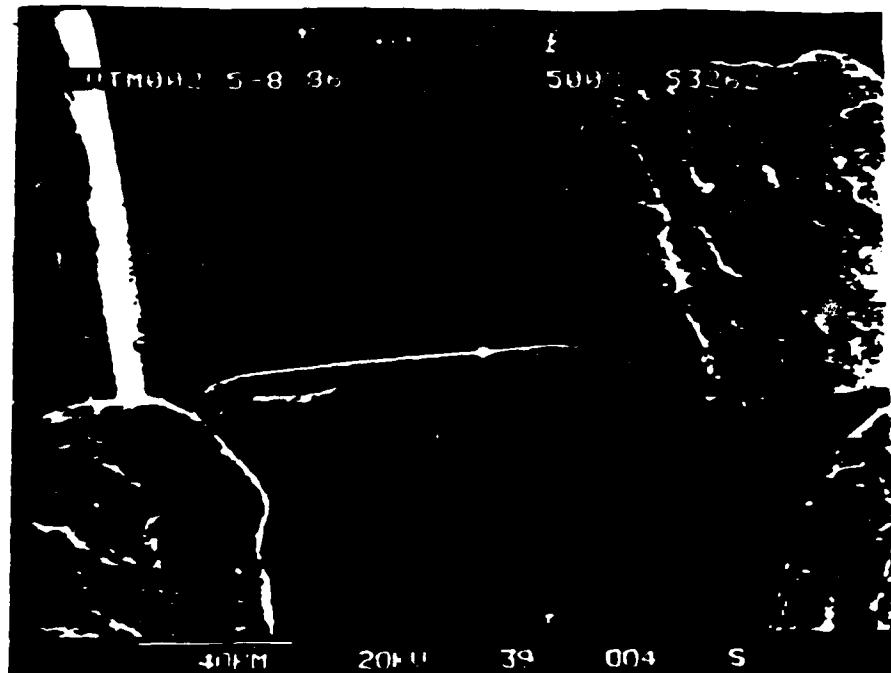


Figure 76. Whisker wire contacting mesa on resonant-tunneling diode chip (Photo courtesy of NASA Jet Propulsion Laboratory).

resistance, the inductance associated with the whisker wire, and the time-varying photoconductive switch,  $R(t)$ . The dc bias to the RTD is connected to the output side of the switch, which is the input side of the device, through a large inductance so that the high frequencies from the switch are incident on the RTD. All the other parasitic circuit elements have been ignored. When an additional current  $\Delta I$  was presented to the RTD as the switch was illuminated and closed, a potential difference developed across the diode, the load line shifted, and the operating point was changed. An observation of the output of the device was made across the load on an ordinary oscilloscope. Although the fast characteristics of the signal were filtered out due to the limited bandwidth of the connectors, cables, and oscilloscope, this observation was still very important because it allowed the experimenter to verify that the excitation beam was correctly aligned on the switch; that is, that there was some sort of an output from the switch transmitted through the

whisker, device, and electrodes. The absence of this output would trigger an investigation into a potential open circuit, such as the whisker contact failing by slipping off the mesa.

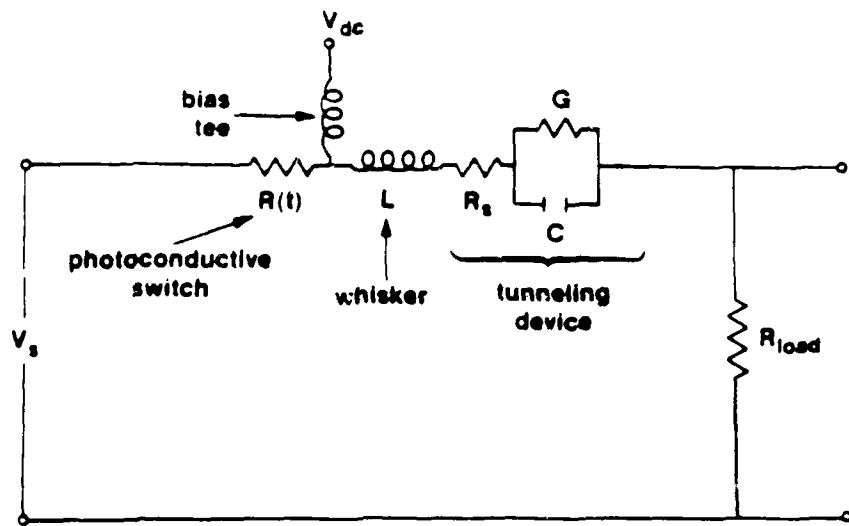


Figure 77: Equivalent circuit for resonant-tunneling diode and test fixture.

#### IV.C.2.. Experimental Observations

As previously mentioned, when the output of the switch was incident upon the RTD, an ac load line like the one for the switching operation in Fig. 71 would move up and to the right an amount proportional to the amplitude of the switch output, all the time remaining parallel to the initial load line. It is assumed here that this ac load line represents the operation of the device at the broad band of high frequencies of interest in this investigation. Since the switching time was to be observed at some point in a waveform that was traversing the RTD and being affected by a wide translation of the load line (not just the part that dropped from a high current to a low current), it was necessary to develop a new way to look at this sort of signal. To explain in another way, if upon the increase in potential due to the switch, the load line ascended so that the output first increased up to the peak of the I-V curve, and then decreased due to the NDR region, then this rapid drop in current would be embedded in the signal appearing at the output of the device. This signal was extracted from the waveforms acquired in this experiment by comparing them with a

reference waveform that was generated by setting the dc bias to zero. When the load line began at the origin, the output of the switch was not great enough to cause it to exceed the I-V curve's peak, and thus it traversed a mostly linear region of the I-V characteristic. When this curve was subtracted from one that had a sudden drop in amplitude, the difference in the two signals gave the evolution of the switching event and the switching time. This (A - B) signal measurement technique allowed the switching to be observed when otherwise it would have been difficult to

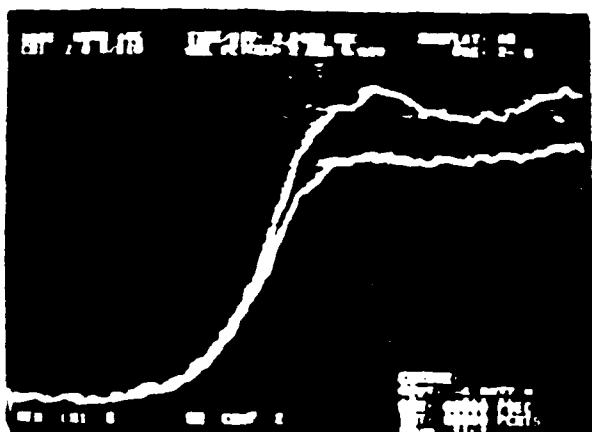


Figure 78. Output waveforms from a resonant tunneling diode, where switching from a high to a low current state has occurred only for the lower-amplitude trace. "A - B" signal measurement would have to be used to extract the rise time of the switching.

resolve in the transmitted waveform. Examples of two waveforms, one causing switching and the other not, are shown in Fig. 78. The upper trace is an output when the load line did not exceed the I-V curve's peak, while the bottom trace occurs when the input pulse to the diode is the same, but the dc bias is such that input has forced the operating point to switch from a high to a low current.

The difference waveforms from four different bias conditions are shown and explained in Fig. 79, where the reference load line follows the I-V curve in the region marked "2" in each of the four cases, even though it is only drawn for the first row. The varying bias is represented by region "1", which follows a different part of the I-V curve in each of the four rows of Fig. 79(a). The I-V characteristic shown in this figure is an approximation of the actual dc device characteristic observed on the curve tracer, where the values have been given previously in this chapter.

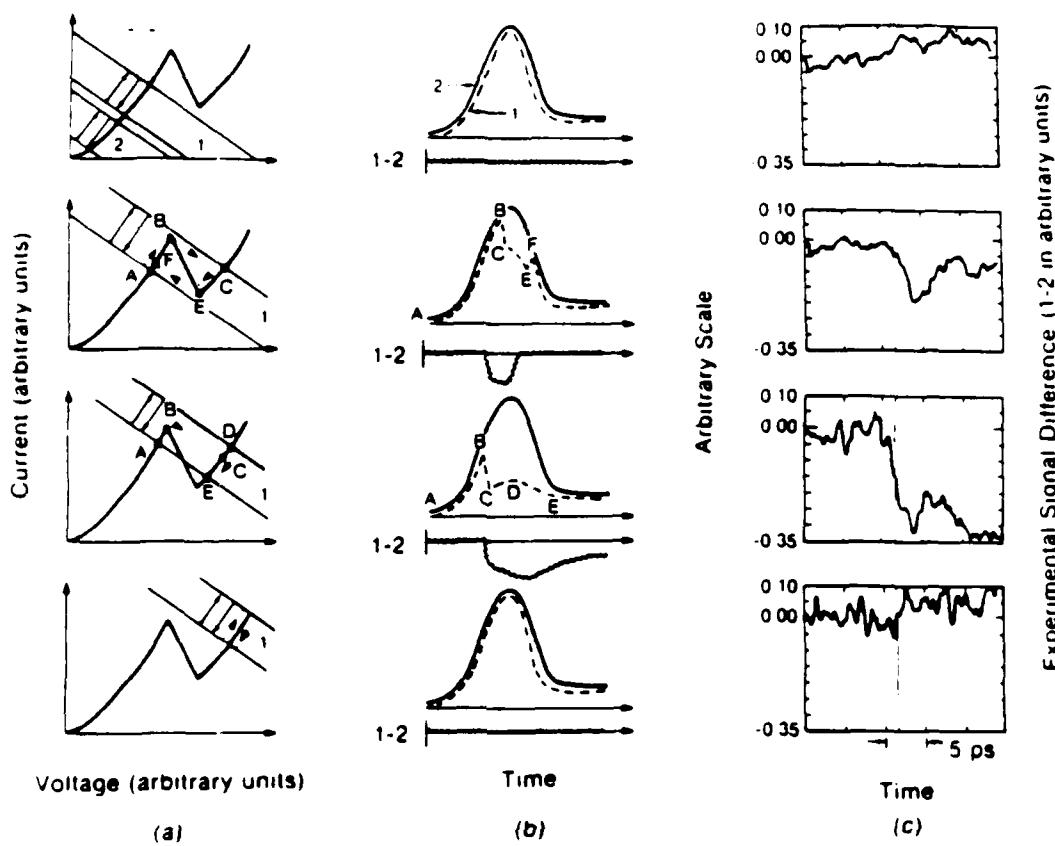


Figure 79. Resonant-tunneling diode switching as a function of bias: (a) I-V curves with load lines; (b) analytic representation of waveforms resulting from movement of load lines; (c) experimental signal difference.

An exaggerated rendition of the behavior with time of the waveforms from the two regions in each of the four cases is schematically given in Fig. 79(b), where the difference has been approximated and sketched below the waveforms. The solid line is for the reference waveform in region 2 in each case, while the dashed line represents the waveforms from region 1. The experimental result of subtracting the two waveforms from regions 1 and 2 is presented in Fig. 79(c).

For the first bias condition, the waveform in region 1 was set so that it would also only traverse a linear regime from the I-V characteristic. The output waveforms were found to be nearly identical to within the noise of the system, and the cancellation between the two was nearly complete as expected. The discrepancies observed were due to noise and to small deviations in the

regions of the I-V curve traversed. The switching process can be observed in the second and third rows of Fig. 79(b) and (c). In the second row, curve 1, the dashed line, followed curve 2 only from points A to B, where it coincided with the rising portion of the I-V curve. When the load line exceeded the current peak of the I-V curve, the diode switched to point C and a rapid drop in current was observed. In the second row, as the input decreased, the optical switch current also decreased so that the load line dropped below point E, and the operating point returned to point F and relaxed to point A, where waveforms 1 and 2 are again nearly identical. This is indicated in the difference waveform (1 - 2) of column (b) and can also be seen in the experimental difference waveform in column (c).

The third row of Fig. 79 displays the same switching action, but with the device attaining a latched state. The operation from points C to D to E was along a path of the I-V curve that approximately cancelled with that from region 2, leading to a long plateau on the difference waveform and a device that had been switched into its lower current state. This was distinctly observed in row 3 of column (c). Since the peak of the I-V characteristic was reached by the load line earlier than in the frame above, the switching took place at an earlier time, as evidenced by the position of the waveforms relative to the thin vertical line of Fig. 79(c) that represents a constant time reference. In row 2, the onset of the switching began at the marker, while in row 3 the switching event was already about halfway completed. For row 3, during the relatively long time of 10 ns between optical pulses, the bias point returns to point A. This is possible because the impedance determining the low-frequency load line is much smaller than the high-frequency impedance, and plus there are negative reflections following the main waveform that would temporarily cause the waveform to descend and the operating point to return to A. In row 4 of this figure, the dc bias point was extended above the area of negative differential resistance, where region 1 traverses virtually the same slope as region 2. The cancellation of the waveforms was again expected and found.

The rise time of the switching event, as most easily measured from the third row Fig. 79(c) was discovered to be 1.9 ps in duration between the 10% and 90% points. The rise time of this

switching event is shown in an expanded view in Fig. 80. This represents the first experimental observation of picosecond bistable operation in a double-barrier diode and is the fastest switching event yet observed for an electronic device. It is also possible that the instrumental response of the electro-optic test fixture has limited the resolution of the measurement, in which case it could be deconvolved with the 1.9-ps rise time to give a potentially shorter

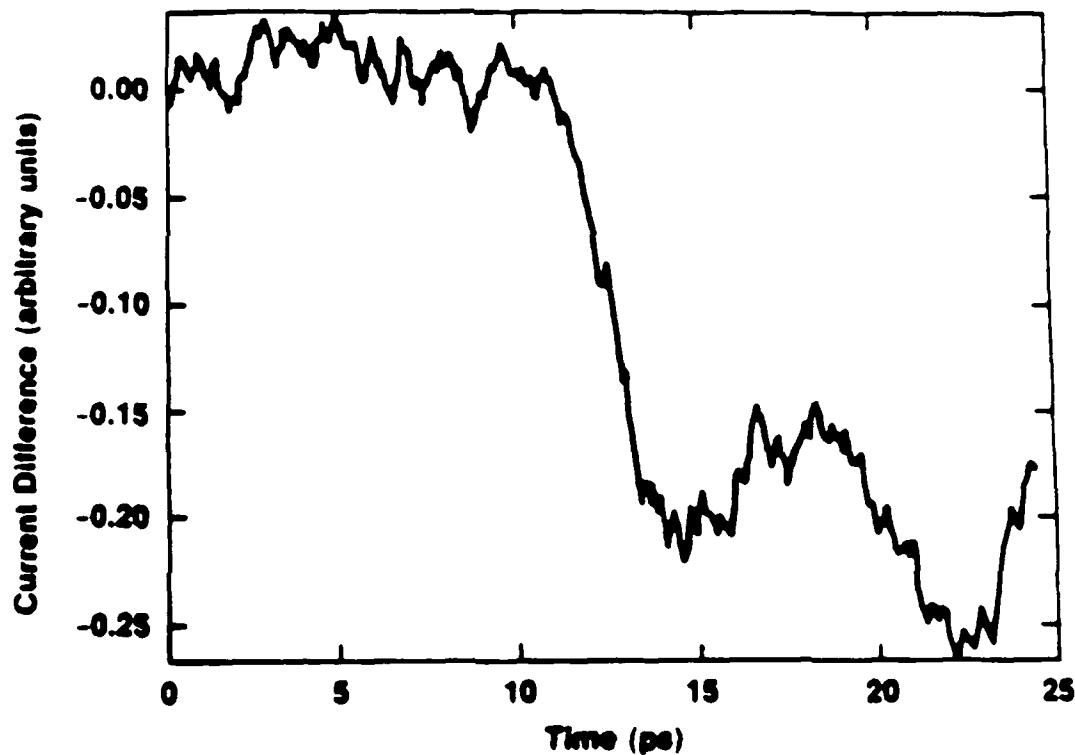


Figure 80. Resonant-tunneling diode switching as a function of time, as it corresponds to the third row of Fig. 79. The rise time of the switching is 1.9 ps.

switching time. Regardless, bearing in mind that Eq. 20 was only approximately valid, the computed switching time of 2.9 ps is in reasonable agreement with the strict measured time of 1.9 ps.

The resonant tunneling diode tested here had a geometry such that  $\Gamma_r$  was about 2 meV. The resulting device response from this RTD could be on the order of a single picosecond. Since the collisionally broadened resonance width would only make this response faster, and the switching experiment is either limited by the circuit or the measurement response, it is impossible

to make any statement about the contribution of either sequential or coherent tunneling from this switching time. However, it would be most interesting to test an RTD having dimensions such that the coherent resonance width was small and the tunneling time was on the order of tens of picoseconds. At room temperature the incoherent resonance width would still cause a fast response so that the switching would be rapid. However, at cryogenic temperatures when the incoherent resonance also became small, the response time should also increase. The unique situation of a decreasing temperature causing an increase in switching time would result. An investigation into this possibility will be undertaken in the future, with the initial work involving the non-trivial problem of contacting the whisker to the diode's mesa in a cryogenic environment.

The investigation undertaken here on the RTD has proven that this kind of experiment on the fast switching time of the device can lead to an increased understanding of its mechanisms of transport. Furthermore, it has shown that the switching speed of the RTD is comparable to those observed in the fastest optical bistable devices.

Well width (Å)	Barrier width (Å)	$\Gamma_r$ (meV)	$\tau_r$ (ps)
50	70	$1.28 \times 10^2$	51.4
50	50	15	4.39
50	30	1.76	0.37
50	20	6.03	0.11

Table 5: Resonance energy line widths and build-up times for AlGaAs/GaAs RTD's of four different barrier dimensions (from 67).

## REFERENCES

1. T. Sizer, II, G. Mourou, and R. R. Rice, "Picosecond Dye Laser Pulses Using a CW Frequency Doubled Nd:YAG as the Pumping Source," *Optics Commun.* **37**, 201 (1981).
2. H. W. Kogelnik, E. P. Ippen, A. Dienes, and C. Shank, "Astigmatically Compensated Cavities for CW Dye Lasers," *IEEE J. Quant. Electron.* **QE-8**, 373 (1972).
3. W. H. Knox, "Femtosecond Optical Pulse Amplification," *IEEE J. Quant. Electron.* **QE-24**, 388 (1988).
4. B. Nikolaus and D. Grischkowsky, "90-fs Tunable Optical Pulses Obtained by Two-Stage Pulse Compression," *Appl. Phys. Lett.* **43**, 228 (1983).
5. R. L. Fork, B. I. Greene, and C. V. Shank, "Generation of Optical Pulses Shorter than 0.1 psec by Colliding Pulse Mode Locking," *Appl. Phys. Lett.* **38**, 671 (1981).
6. J. Valdmanis, R. L. Fork, and J. P. Gordon, "Generation of Optical Pulses as Short as 27 Femtoseconds Directly from a Laser Balancing Self-Phase Modulation, Group Velocity Dispersion, Saturable Absorption, and Saturable Gain," *Opt. Lett.* **10**, 131 (1985).
7. R. L. Fork, C. V. Shank, and R. T. Yen, "Amplification of 70-fs Optical Pulses to Gigawatt Powers," *Appl. Phys. Lett.* **41**, 223 (1982).
8. W. H. Knox, M. C. Downer, R. L. Fork, and C. V. Shank, "Amplified Femtosecond Optical Pulses and Continuum Generation at 5-kHz Repetition Rate," *Opt. Lett.* **9**, 552 (1984).
9. I. N. Duling, III, T. Norris, T. Sizer, II, P. Bado, and G. A. Mourou, "Kilohertz Synchronous Amplification of 85-Femtosecond Optical Pulses," *J. Opt. Soc. Am. B* **2**, 616 (1985).
10. T. Norris, T. Sizer, II, and G. Mourou, "Generation of 85-fsec Pulses by Synchronous Pumping of a Colliding-Pulse Mode-Locked Dye Laser," *J. Opt. Soc. Am. B* **2**, 613 (1985).
11. T. Norris, I. N. Duling, III, M. Pessot, T. Sizer, II, J. Dawes, and G. A. Mourou, "Generation of Microjoule 65-fsec Pulses at High Repetition Rate," in Digest of Conference on Lasers and Electro-Optics (Optical Society of America, Washington, D.C., 1985), paper TUE4.
12. A. E. Siegman, "An Antiresonant Ring Interferometer for Coupled Laser Cavities, Laser Output Coupling, Mode Locking, and Cavity Dumping," *IEEE J. Quant. Electron.* **QE-9**, 247 (1973).
13. A. E. Siegman, "Passive Mode Locking Using an Antiresonant-Ring Laser Cavity," *Opt. Lett.* **6**, 334 (1981).
14. H. Vanherzeele, J. L. Van Eck, and A. E. Siegman, "Colliding Pulse Mode Locking of a Nd:YAG Laser with an Antiresonant Ring Structure," *Appl. Opt.* **20**, 3484 (1981).
15. J. M. Buchert, D. K. Basa, L. Tzu and R. R. Alfano, "Colliding Pulse Mode Locking for an Antiresonant Cavity of a Nd:Glass Laser," *J. Appl. Phys.* **55**, 683 (1984).
16. S. De Silvestri, P. Laporta, and O. Svelto, "Effects of Cavity Dispersion on Femtosecond Mode-Locked Dye Lasers," in Ultrafast Phenomena IV, edited by D. H. Auston and K. B. Eisenthal, (Springer-Verlag, Berlin, 1984), p. 23.
17. R. L. Fork, O. E. Martinez, and J. P. Gordon, "Negative Dispersion Using Pairs of Prisms," *Opt. Lett.* **9**, 150 (1984).
18. T. Sizer, II, J. D. Kafka, A. Krisiloff, and G. Mourou, "Generation and Amplification of Sub-Picosecond Pulses Using a Frequency-Doubled Neodymium YAG Pumping Source," *Opt. Commun.* **39**, 259 (1981).
19. R. R. Alfano and S. L. Shapiro, "Emission in the Region 4000-7000 Å via Four-Photon Coupling in Glass," *Phys. Rev. Lett.* **24**, 584 (1970).
20. R. L. Fork, C. V. Shank, C. Hirlmann, R. Yen, and W. J. Tomlinson, "Femtosecond White-Light Continuum Pulses," *Optics Lett.* **8**, 1 (1983).

21. J. T. Manassah, M. A. Mustafa, R. R. Alfano, and P. P. Ho, "Spectral Extent and Pulse Shape of the Supercontinuum for Ultrashort Laser Pulse," *IEEE J. Quant. Electron. QE-22*, 197 (1986), and references therein.
22. J. A. Valdmanis, G. A. Mourou, and C. W. Gabel, *IEEE J. Quant. Electron. QE-19*, 664 (1983).
23. J. A. Valdmanis, "Subpicosecond Electro-Optic Sampling," Ph.D. Thesis, University of Rochester, 1983.
24. G. A. Mourou and K. W. Meyer, *Appl. Phys. Lett.* **45**, 492 (1984).
25. D. H. Auston, K. P. Cheung, J. A. Valdmanis, and D. A. Kleinman, *Phys. Rev. Lett.* **53**, 1555 (1984).
26. K. J. Weingarten, M. J. W. Rodwell, H. K. Heinrich, B. H. Kolner, and D. M. Bloom, *Electron. Lett.* **21**, 765 (1985).
27. R. K. Jain, D. E. Snyder, and K. Stenersen, *IEEE Electron. Dev. Lett.* **EDL-5**, 371 (1984).
28. K. E. Meyer and G. A. Mourou, *Electron. Lett.* **21**, 568 (1985).
29. J. A. Valdmanis, Workshop on Interconnections of High Speed and High Frequency Devices and Systems, sponsored by SPIE, Newport Beach, California, March 1988 (to be published).
30. J. Nees and G. Mourou, *Electron. Lett.* **22**, 918 (1986).
31. M. A. Osman and D. K. Ferry, *J. Appl. Phys.* **61**, 5330 (1987).
32. K. Brennan and K. Hess, *Phys. Rev. B* **29**, 5581 (1984).
33. Robert Grondin, private communication.
34. D. Jones and H. D. Rees, *J. Phys. C* **6**, 1781 (1973).
35. R. O. Grondin and M. J. Kann, Proceedings of the 1987 Hot Electron Conference, Boston, Massachusetts, July 1987 (to be published).
36. D. H. Auston, *IEEE J. Quant. Electron. QE-19*, 639 (1983).
37. K. E. Meyer, M. Pessot, G. Mourou, R. L. Grondin, and S. Chamoun, submitted to *Physical Review Letters*.
38. W. H. Knox, C. Hirlmann, D. A. B. Miller, J. Shah, D. S. Chemla, and C. V. Shank, *Phys. Rev. Lett.* **56**, 1191 (1986).
39. J. Collet, J. L. Oudar, and T. Amand, *Phys. Rev. B* **34**, 5443 (1986).
40. D. W. Bailey, M. A. Artaki, C. J. Stanton, and K. Hess, *J. Appl. Phys.* **62**, 4638 (1987).
41. W. H. Knox, M. C. Downer, R. L. Fork, and C. V. Shank, *Opt. Lett.* **9**, 552 (1984).
42. J. D. Kafka and T. Baer, *Opt. Lett.* **12**, 401 (1987).
43. C. J. Wei, H. J. Klein, and H. Beneking, *IEEE Trans. Electron. Dev. ED-29*, 1442 (1982).
44. P. Jeppesen and B. I. Jeppson, *IEEE Trans. Electron. Dev. ED-20*, 371 (1973).
45. L. D. Landau and E. M. Lifshitz, Quantum Mechanics, Non-Relativistic Theory, Third Edition (Pergamon Press, New York, 1977), pp. 178-181.
46. M. Tsuchiya, T. Matsusue, and H. Sakaki, "Tunneling Escape Rate of Electrons from Quantum Well in Double-Barrier Heterostructures," *Phys. Rev. Lett.* **59**, 2356 (1987).
47. T. C. L. G. Sollner, P. E. Tannenwald, D. D. Peck, and W. D. Goodhue, "Quantum Well Oscillators," *Appl. Phys. Lett.* **45**, 1319 (1984).
48. E. R. Brown, T. C. L. G. Sollner, W. D. Goodhue, and C. D. Parker, "Millimeter-Band Oscillations Based on Resonant Tunneling in a Double-Barrier Diode at Room Temperature," *Appl. Phys. Lett.* **50**, 83 (1987).
49. T. C. L. G. Sollner, E. R. Brown, W. D. Goodhue and H. Q. Le, "Observation of Millimeter-Wave Oscillations from Resonant Tunneling Diodes and Some Theoretical Consideration of Ultimate Frequency Limits," *Appl. Phys. Lett.* **50**, 332 (1987).
50. T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker and D. D. Peck, "Resonant Tunneling Through Quantum Wells at Frequencies up to 2.5 THz," *Appl. Phys. Lett.* **43**, 588 (1983).
51. B. Ricco and M. Ya. Azbel, "Physics of Resonant Tunneling. The One-Dimensional Double-Barrier Case," *Phys. Rev. B* **29**, 1970 (1984).

52. S. Luryi, "Frequency Limit of Double-Barrier Resonant-Tunneling Oscillators," *Appl. Phys. Rev.* **47**, 490 (1985).
53. A. D. Stone and P. A. Lee, "Effect of Inelastic Processes on Resonant Tunneling in One Dimension," *Phys. Rev. Lett.* **54**, 1196 (1985).
54. D. D. Coon and H. C. Liu, "Time-Dependent Quantum-Well and Finite-Superlattice Tunneling," *J. Appl. Phys.* **58**, 2230 (1985).
55. D. D. Coon and H. C. Liu, "Frequency Limit of Double Barrier Resonant Tunneling Oscillators," *Appl. Phys. Lett.* **49**, 94 (1986).
56. W. R. Frensel, "Transport Response of a Tunneling Device Obtained from the Wigner Function," *Phys. Rev. Lett.* **57**, 2853 (1986).
57. H. C. Liu, "Tunneling Time Through Heterojunction Double Barrier Diodes," *Superlattices and Microstructures* **3**, 379 (1987).
58. T. Weil and B. Vinter, "Equivalence Between Resonant Tunneling and Sequential Tunneling in Double-Barrier Diodes," *Appl. Phys. Lett.* **50**, 1281 (1987).
59. M. Jonson and A. Grincwajg, "Effect on Inelastic Scattering on Resonant and Sequential Tunneling in Double Barrier Heterostructures," *Appl. Phys. Lett.* **51**, 1729 (1988).
60. H. C. Liu, "Time-Dependent Approach to Double-Barrier Quantum Well Oscillators," *Appl. Phys. Lett.* **52**, 453 (1988).
61. J. D. Bruno, T. B. Bahder, and C. A. Morrison, "Limiting Response Time of Double-Barrier Resonant Tunneling Structures," *Phys. Rev. B* **37**, 7098 (1988).
62. P. J. Price, "Theory of Resonant Tunneling in Heterostructures," *Phys. Rev. B* **38**, 1994 (1988).
63. H. Guo, K. Diff, G. Neofotistos, and J. D. Gunton, "Time-Dependent Investigation of the Resonant Tunneling in a Double-Barrier Quantum Well," *Appl. Phys. Lett.* **53**, 131 (1988).
64. J. F. Whitaker, G. A. Mourou, T. C. L. G. Sollner, and W. D. Goodhue, "Picosecond Switching Time Measurement of a Resonant Tunneling Diode," *Appl. Phys. Lett.* **53**, 385 (1988).
65. R. Gupta and B. K. Ridley, "The Effect of Level Broadening on the Tunneling of Electrons Through Semiconductor Double-Barrier Quantum-Well Structures," *J. Appl. Phys.* **64**, 3089 (1988).
66. S. C. Kan and A. Yariv, "The Ultimate Frequency Response and Time Evolution of Resonant Tunneling in Double-Barrier Structures," *J. Appl. Phys.* **64**, 3098 (1988).
67. F. Capasso, K. Mohammed, and A. Y. Cho, "Resonant Tunneling Through Double Barriers, Perpendicular Quantum Transport Phenomena in Superlattices, and Their Device Applications," *IEEE J. Quant. Electron.* **QE-22**, 1853 (1986).
68. The sample growth and processing were done at Cornell University by Xiao Song, with advice and help from Bill Schaff.
69. J. Shah, A. Pinczuk, A. C. Gossard, and W. Wiegmann, "Energy-Loss Rates for Hot Electrons and Holes in GaAs Quantum Wells," *Phys. Rev. Lett.* **54**, 2045 (1985).
70. J. F. Ryan, R. A. Taylor, A. J. Turberfield, A. Maciel, J. M. Worlock, A. C. Gossard, and W. Wiegmann, "Time-Resolved Photoluminescence of Two-Dimensional Hot Carriers in GaAs-AlGaAs Heterostructures," *Phys. Rev. Lett.* **53**, 1841 (1984).
71. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "Electric Field Dependence of Optical Absorption Near the Band Gap of Quantum Well Structures," *Phys. Rev. B* **32**, 1043 (1985).
72. H. -J. Polland, L. Schultheiss, J. Kuhl, E. O. Gobel, and C. W. Tu, "Lifetime Enhancement of Two-Dimensional Excitons by the Quantum-Confinement Stark Effect," *Phys. Rev. Lett.* **55**, 2610 (1985).
73. E. J. Austin and M. Jaros, "Electric Field Induced Shifts and Lifetimes in GaAs-GaAlAs Quantum Wells," *Appl. Phys. Lett.* **47**, 274 (1985).
74. S. M. Sze, *Physics of Semiconductor Devices*, Second Edition, (John Wiley and Sons, New York, 1981).

75. X. J. Song, private communication, 1988.
76. X. J. Song, private communication, 1988.
77. D. A. B. Miller, *et al.*, *op. cit.*, Figure 2.
78. H. Haug and S. Scmitt-Rink, "Electron Theory of the Optical Properties of Laser-Excited Semiconductors," *Prog. Quantum Electron.* **9**, 3 (1984).
79. G. D. Sanders and Yia-Chung Chang, "Theory of Photoabsorption in Modulation-Doped Semiconductor Quantum Wells," *Phys. Rev. B* **35**, 1300 (1987).
80. H. Haug and S. Scmitt-Rink, "Basic Mechanisms of the Optical Nonlinearities of Semiconductors Near the Band Edge," *J. Opt. Soc. Am. B* **2**, 1135 (1985).
81. G. E. W. Bauer and T. Ando, "Electronic Structure of Free Carriers in Quantum Wells Calculated by Density-Functional Theory," *Phys. Rev. B* **34**, 1300 (1986).
82. G. Livescu, D. A. B. Miller, D. S. Chemla, M. Ramaswamy, T. Y. Chang, N. Sauer, A. C. Gossard, and J. H. English, "Free Carrier and Many-Body Effects in Absorption Spectra of Modulation-Doped Quantum Wells," *IEEE J. Quantum Electron.* **QE-24**, 1677 (1988).
83. G. Wicks, private communication, 1986.
84. B. Vinter, private communication, 1988.
85. E. J. Austin and M. Jaros, "Electric Field Induced Shifts and Lifetimes in GaAs-GaAlAs Quantum Wells," *Appl. Phys. Lett.* **47**, 274 (1985).
86. S. M. Sze, Physics of Semiconductor Devices, Second Edition, (John Wiley and Sons, New York, 1981), p. 291.
87. T. Tada, A. Yamaguchi, T. Ninomiya, H. Uchiki, T. Kobayashi, and T. Yao, "Tunneling Process in AlAs/GaAs Double Quantum Wells Studied by Photoluminescence," *J. Appl. Phys.* **63**, 5491 (1988).
88. N. Vodjdani, private communication, 1988.
89. H. -J Polland, W.W. Ruhle, J. Kuhl, K. Ploog, K. Fujiwara, and T. Nakayama, "Nonequilibrium Cooling of Thermalized Electrons and Holes in GaAs/AlGaAs Quantum Wells," *Phys. Rev. B* **35**, 8273 (1987).
90. D. Park, Introduction to the Quantum Theory, (McGraw-Hill, New York, 1974) Chapter 4.
91. C. Weisbuch, "Fundamental Properties of III-V Semiconductor Two-Dimensional Quantized Structures: The Basis for Optical and Electronic Device Applications," in Semiconductors and Semimetals, R. Dingle, Ed., (Academic Press, San Diego, 1987) Chapter 1.
92. B. Ricco and M. Ya. Azbel, "Physics of Resonant Tunneling. The One Dimensional Double-Barrier Case," *Phys. Rev. B* **29**, 1970-1981 (1984).
93. F. Capasso, K. Mohammed, and A. Y. Cho, "Resonant Tunneling Through Double Barriers. Perpendicular Quantum Transport Phenomena in Superlattices, and Their Device Applications," *IEEE J. Quantum Electron.* **QE-22**, 1853-1869 (1986).
94. H. C. Liu and D. D. Coon, "Heterojunction Double-Barrier Diodes for Logic Applications," *Appl. Phys. Lett.* **50**, 1246-1248 (1987).
95. E. R. Brown, T. C. L. G. Sollner, W. D. Goodhue, and C. D. Parker, "Millimeter-band Oscillations Based on Resonant Tunneling in a Double-Barrier Diode at Room Temperature," *Appl. Phys. Lett.* **50**, 83-85 (1987).
96. California Institute of Technology, Jet Propulsion Laboratory, Procedure No. EP513877, "Detail Procedure for 205 GHz Mixer Assembly Microwave Limb Sounder," Pasadena, California, (1986).